Microbial ecology of the sheep mammary gland

Emma Monaghan

BSc. & MSc.

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University of Warwick

School of Life Sciences

September 2014

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would firstly like to thank the Biotechnology and Biological Sciences Research Council (BBSRC), English Beef and Lamb Executive (EBLEX) and the BioSciences Knowledge Transfer Network (KTN) for funding this study. I would also like to thank the industrial partner Quality Milk Management Services Ltd (QMMS, Somerset, UK) and the School of Life Sciences at the University of Warwick for hosting me from 2010-2014.

My sincere thanks to my supervisors: Professor Laura Green, Dr Kevin Purdy and Dr Andrew Bradley. The support, guidance and encouragement they have provided has made this thesis possible. I cannot thank them enough for everything they have taught me which has made me a more confident and independent scientist.

Thanks must also go to Dr Selene Huntley for sample collection, Dr Simon Williams for bioinformatic support and Dr Ed Smith for general advice and help throughout my PhD. Thanks to both past and present PhD students and post-docs from both Laura and Kevin's group over the past four years. Working with you all has been a pleasure and I have enjoyed it immensely.

Last but not least, thanks to my family for your support and understanding, I know you sometimes have no idea what I do, but I hope you are proud of my achievements anyway!

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself under the supervision of Professor Laura Green, Dr Kevin Purdy and Dr Andrew Bradley and has not been submitted in any previous application for any degree.

Summary

Intramammary infections (IMI) in sheep have a major economic impact through reduced milk production, premature culling and even death of ewes. This study hypothesizes that the sheep mammary gland could host a microbiome with certain members affecting SCC.

Previous studies have been cross-sectional, using only one sample per subject and not conducted in sheep. This limits understanding causality; that is, how infection develops and what triggers development of disease.

A longitudinal study of 30 sheep, each with two mammary gland halves, collected over 8 weeks, provided 379 milk samples and data on ewe parity and milk SCC. DNA was extracted from milk samples and processed using a bacterial 16S rRNA gene targeted PCR. Bacterial community diversity was visualised using denaturing gradient gel electrophoresis (DGGE).

DGGE fingerprints were analysed in a mixed effects regression model to identify associations between individual DGGE bands and changes in SCC. Those bands associated with SCC were sequenced. *Corynebacterium efficiens*, *Psychrobacter maritimus*, *Streptococcus uberis*, *Burkholderia cepacia*, *Fusobacterium necrophorum*, *Trueperella pyogenes*, *Pseudomonas chlororaphis* and *Psychrobacter faecalis* were significantly associated with a higher SCC. *Achromobacter xylosoxidans*, *Nocardia globerula* or *Rhodococcus qingshengii*, *Atopostipes suicloacalis*, *Mannheimia haemolytica*, *Jeotgalicoccus psychrophilus* and *Sharpea azabuensis* were significantly associated with a lower SCC.

A protocol to analyse all study samples using Illumina MiSeq sequencing was developed to elucidate the complex interactions between the sheep mammary gland microbiome and SCC.

The DGGE and MiSeq results show a persistent community has been detected over time, with similarities and differences by mammary gland half, lactation and age. Associations between individual bacterial species and SCC were identified through mixed effect modelling. The DGGE results were comparable to the MiSeq results from 5 sheep.

Analysis of all 379 samples by MiSeq sequencing and mixed effects models will be used to directly test the study hypotheses.

List of abbreviations

ANOSIM	Analysis of similarity
A.U.	Arbitrary unit
BLAST	Basic local alignment search tool
BMSCC	Bulk milk somatic cell count
Bp	Base pairs
BHI agar	Brain heart infusion agar
BSA	Bovine serum albumin
Cells/ml	cells per millilitre
Cfu/ml	colony forming units per millilitre
CNS	Coagulase negative staphylococci
DGGE	Denaturing gradient gel electrophoresis
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
G -ve	Gram negative
G +ve	Gram positive
IMI	Intramammary infection
HTP	Hydroxyapaptite
LB agar	Luria Bertani agar
MDS	Multi-dimensional scaling plot
MG	Mammary gland
M. haemolytica	*Mannheimia haemolytica*
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
QMMS	Quality Milk Management Services Ltd
ReLIC	Read length incremental clustering
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
SCC	Somatic cell count
SDS	Sodium dodecyl sulfate
S. aureus	*Staphylococcus aureus*
S. hyicus	*Staphylococcus hyicus*
Strep. agalactiae	*Streptococcus agalactiae*
Strep. dysgalactiae	*Streptococcus dysgalactiae*
Strep. uberis	*Streptococcus uberis*
UPGMA	Unweighted pair group method with arithmetic mean
UV	Ultraviolet

Chapter 1 : Introduction

1.1 Study of the microbiome

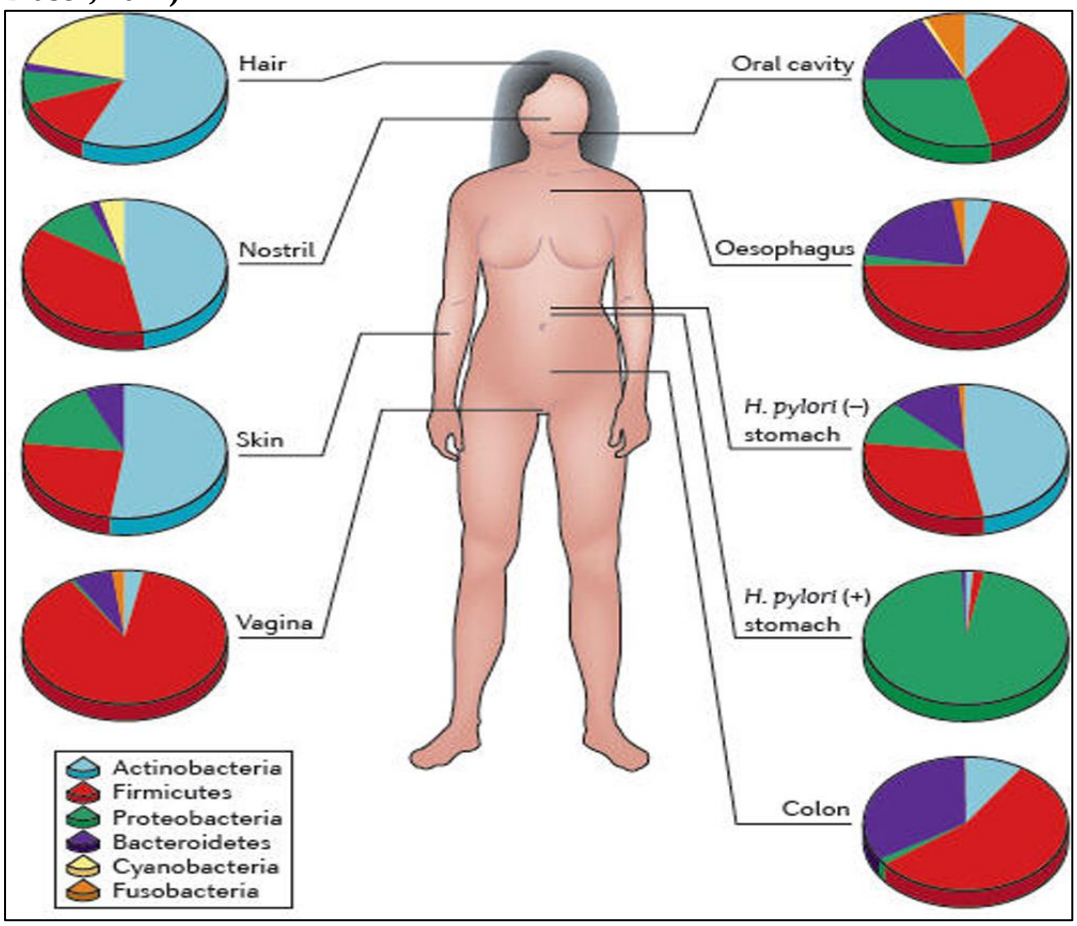
Microbial communities are defined as multi-species assemblages in which organisms live and interact in a shared environment. Such communities form from populations of bacteria conducting interdependent physiological processes (Davey and O'Toole, 2000). The term 'microbiome' refers to the totality of microbes, their genetic information and the milieu in which they interact. Microbiomes typically consist of environmental or biological niches containing complex communities of microbes (Cho and Blaser, 2012). Most host-associated microbes are difficult to culture in their entirety because laboratory conditions are not conducive to growth of some organisms. Advances in microbial ecology have revolutionised our understanding of microbial communities and their roles, with large-scale studies on host-microbe and microbe-microbe interactions now possible (Proctor, 2011). This had led to an improved understanding of the essential functions microbial communities play in the maintenance of health in higher-order organisms (Morgan *et al.*, 2013; Proctor, 2011).

A key example that demonstrates the important role of microbial communities is the human microbiota. The term 'microbiota' refers to the microbial organisms that constitute the microbiome. The microbiota can vary substantially according to host niche, environmental site and between health status (Cho and Blaser, 2012). The Human Microbiome Project generated a 16S RNA metagenomic data set of over 35 billion reads from 690 samples from over 200 subjects across 15-18 body sites (Methe *et al.*, 2012; Turnbaugh *et al.*, 2007). Such large-scale studies provide a preliminary understanding of the biology of the human microbiome and its possible role in health and disease.

The human microbiome consists of communities across several anatomical sites including the skin, oral cavity, breast milk, vagina and gastro-intestinal tract as illustrated in Figure 1-1. The development of a microbial community is influenced by the role the community plays in the host organism in addition to influencing factors from the external environment and other associated microbial populations. Membership of these communities is diverse, one survey of the skin microbiota from 10 volunteers identified 19 bacterial phyla and 205 genera (Grice *et al.*, 2008), with even the low pH environment of the stomach averaging 10^3 bacteria g^{-1} (Holzapfel *et al.*, 1998). These communities have co-evolved with the human host to form an

essential part of the genetic composition that is vital in maintaining health (Costello *et al.*, 2009; Turnbaugh *et al.*, 2007).

Figure 1-1: Differences in human microbiome according to anatomical site (Cho and Blaser, 2012)



The development of a microbial community is influenced by the role the community plays in the host organism in addition to influencing factors from the external environment and other associated microbial populations. Recent research into the human milk microbiome where 3 consecutive samples were taken, has provided evidence of a bacterial community that is probably not the result of contamination (Cabrera-Rubio *et al.*, 2012; Fernández *et al.*, 2012; Hunt *et al.*, 2011).

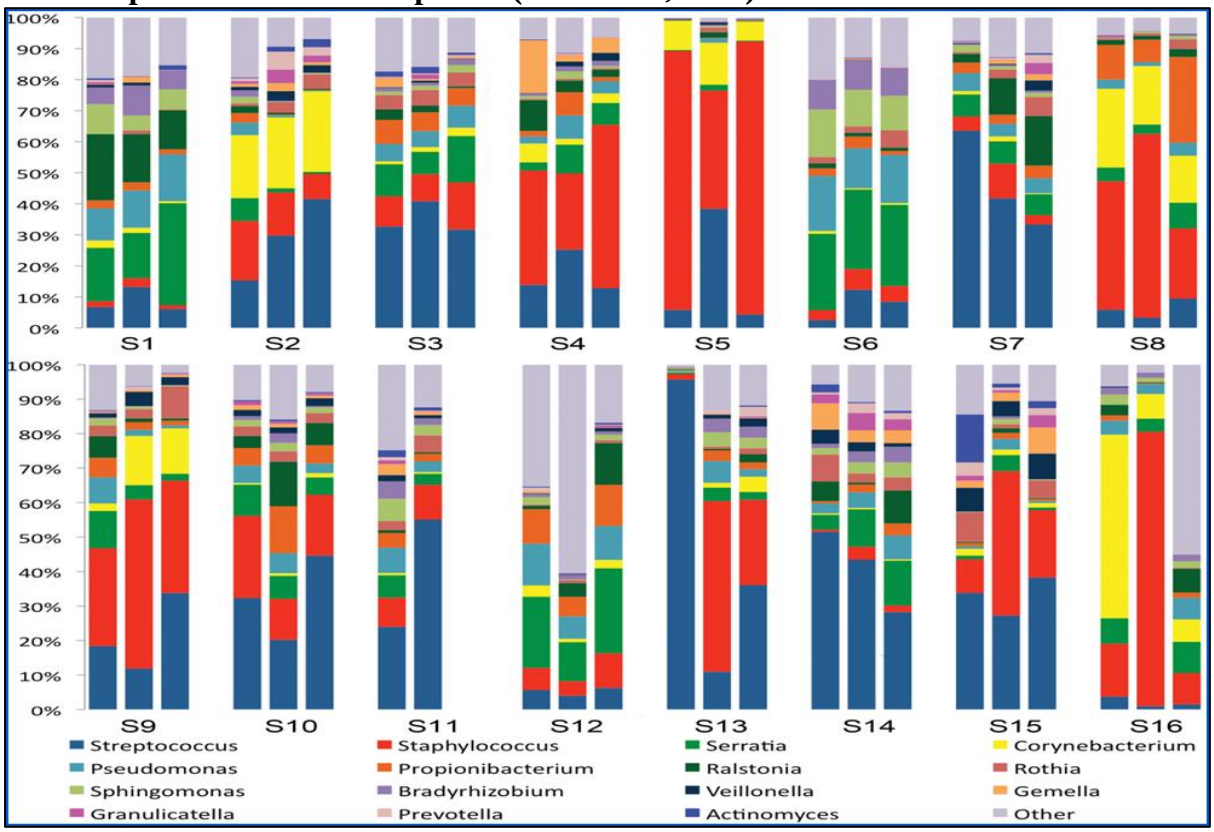
Hunt *et al.*, (2011) characterised bacterial communities in milk samples collected from 16 women at 3 time points over 4 weeks and found a diverse community of between 100-600 OTUs per woman, dominated by *Staphyloccocus*, *Streptococcus*, *Serratia* and *Corynebacteria*. Phylotypes such as *Staphylococcus* and *Corynebacteria* are typically present on human skin (Grice *et al.*, 2008). However, despite sharing several bacterial phylotypes,

major differences from the skin microbiota were identified, suggesting that bacterial communities in milk are not simply the result of skin contamination.

Figure 1-2 shows the 15 most abundant bacterial genera identified in the assessment of human breast milk by Hunt *et al.*, (2011). After the 4 most abundant bacterial genera, 8 other genera represented ≥1% of the communities observed across milk samples. Interestingly, the stability and composition of individual subjects was variable. For example, in subject 5, milk samples were dominated by *Staphylococcus*, whereas in subject 3, *Staphylococcus* was only a minor contributor to the community. Some milk samples were consistent over time e.g. subjects 1 and 3, whilst for others the relative abundance of the detected genera varied over time e.g. subjects 13 and 16.

Despite the evidence of individual bacterial communities over time, a "core" microbiome of 9 OTUs were present in every milk sample, representing 50% of the bacterial abundance in communities within women. This does indicate that 50% of the community was not conserved across women, but these results contrast with studies of microbiomes at other anatomical sites as such as the gut and vagina, where no highly abundant OTUs were shared between individuals (Ravel *et al.*, 2011; Turnbaugh *et al.*, 2007).

Figure 1-2: 15 most abundant bacterial genera in 47 milk samples taken from 16 women at 3 time points over a 4 week period (Hunt *et al.*, 2011)



However, microbiomes at different anatomical sites are not isolated, but rather a network of inter-related communities that experience change. Costello *et al.,* (2009) analysed microbiota samples from several body sites including to up 18 skin locations and found members of 22 bacterial phyla, with most sequences relating to 4 phyla; *Actinobacteria*, *Firmicutes*, *Proteobacteria* and *Bacteroidetes*. Each habitat had its own characteristic microbiota as well as a relatively stable set of abundant taxa across individuals and over time. It is therefore likely that milk communities in animals such as cows and sheep are no different, with exposure and interaction with other microbial populations and environmental influences resulting in complex communities that change over time.

Several studies have investigated the milk microbiome of dairy cattle using high-throughput sequencing (Bhatt *et al.*, 2012; Kuehn *et al.*, 2013; Oikonomou *et al.*, 2014; Oikonomou *et al.*, 2012).

Kuehn *et al.,* (2013) identified microbial communities in culture-negative milk samples, highlighting the power of a culture-independent approach. The same study revealed significant differences in the microbiota of milk samples from diseased and healthy MG quarters. Higher abundances of *Brevundimonas*, *Burkholderia*, *Sphingomonas* and *Stenotrophomonas* were found in diseased samples. However, *Sphingomonas* spp. were also predominant in culture-negative diseased milk samples in a study by Oikonomou *et al.,* (2014). In healthy milk samples, *Pseudomonas*, *Psychrobacter* and *Ralstonia* were most prevalent. It could be speculated from this that *Sphingomonas* spp. might be associated with healthy intramammary communities. The difference in microbiota according to disease state suggests that differences between multiple milk samples from the same animal could be related to disease state as opposed to variability between individual animals. However, variation in the microbial community according to mammary gland quarter as well as animal-specific microbiomes were also identified.

Kuehn *et al.,* (2013), found *Staphlococcus* and *Corynebacterium* in both diseased and healthy samples as did Hunt *et al.,* (2011) in human milk. These bacteria have also been associated with the teat skin (Verdier-Metz *et al.,* 2009), suggesting a potential overlap in the microbiota of these environments. However, given that the milk microbiome in humans is compositionally distinct from the skin microbiome, there is no reason to think that the milk

microbiome of animals is not also ecologically distinct, with some phylotypes potentially shared with related communities.

Interestingly, the bacterial species Kuehn *et al.,* (2013) found most commonly in healthy milk samples have not been linked to causing intramammary disease. This could suggest that these organisms are commensal, or play a stabilising and/or protective role in the mammary gland. Supporting this, Oikonomou *et al.,* (2014) found *Propionibacterium acnes* most prevalent in healthy milk samples and Shu *et al.,* (2013) reported that fermentation of *Propionibacterium acnes* may play a role in human innate immunity against *Staphylococcus aureus*.

A study by Oikonomou *et al.,* (2012) also found differences in microbial community composition between healthy and diseased milk samples in cattle. For example, the anaerobe *Fusobacterium necrophorum* was highly prevalent in diseased milk samples that were diagnosed as intramammary infections caused by *Trueperella pyogenes* and in low prevalence in healthy milk samples in a second study (Oikonomou *et al.,* 2014). Anaerobes were also identified in a study of subclinical infection by Bhatt *et al.,* (2012). A synergistic relationship between the anaerobes *Fusobacterium necrophorum* and *Trueperella pyogenes* in the development of intramammary infection has been reported previously (Pyorala *et al.,* 1992). It could be postulated that the role of anaerobes has been underestimated due to the use of aerobic bacterial culture as the gold standard in identifying the causative agents of intramammary infections.

However, the communities in normal and diseased milk samples in the Oikonomou *et al.,* (2012) study were not entirely separate as *Streptococcus* spp. were prevalent in all groups of milk samples. Hunt *et al.,* (2011) identified *Streptococcus* spp. as one of the most prevalent bacteria in healthy human breast milk. As *Streptococcus* spp. such as *Streptococcus uberis* have been associated with IMI in dairy cattle (Hillerton and Berry, 2005), such findings could suggest that bacterial pathogens can be present when there is no clinical disease and so other factors such as relative abundance of different community members and synergistic interactions between microorganisms may play a role in disease progression.

In relation to this, Oikonomou *et al.,* (2012) often found more than one bacterial pathogen in diseased milk samples and Bhatt *et al.,* (2012) suggested that subclinical disease is not caused by a single pathogenic species of bacteria, but rather a blend of several microbes. Bhatt *et al.,* (2012) also found that *Firmicutes* and *Proteobacteria* were the main phyla in subclinically diseased milk samples from 3 breeds of dairy cattle. However, the abundance of these phyla

was different in different breeds of cattle. These two phyla were also found in the study by Costello *et al.,* (2009) in multiple microbial communities across various sites of the human body. This suggests that some microorganisms are common to multiple sites in both humans and animals, highlighting the influence that the interactions between different microbial communities in close proximity to one another may have on the health state of the host.

A study by Oikonomou *et al.,* in (2014) also found differences in microbial community between milk samples from normal and diseased dairy cattle as well as differences according to farm. In normal quarters, *Faecalibacterium* spp., unclassified *Lachnospiraceae*, *Propionibacterium* spp. and *Aeribacillus* spp. were most prevalent. As found in other studies (Hunt *et al.*, 2011; Oikonomou *et al.*, 2012), *Staphylococcus* and *Streptococcus* were most prevalent in all milk samples, regardless of disease status.

However, differences between human and dairy cattle milk samples have been identified. For example, Hunt *et al.,* (2011) found very few sequences for *Lactobacillus* whereas Oikonomou *et al.,* (2014) found *Lactobacillus* was prevalent in healthy cow milk. Some *Lactobacillus* spp. have been reported as capable of inhibiting intramammary pathogens including *Escherichia coli* (Jara *et al.*, 2011), offering a potential hypothesis for its presence in healthy milk. Differences in milk communities amongst different hosts whether human or animal is expected, as a range of factors including environment, diet, lifestyle and genetic influences as well as differences between study designs and methodologies may affect the detected microbiota composition.

Oikonomou *et al.,* (2014) also points to the presence of a commensal microbial flora previously suggested by Kuehn *et al.,* (2013) who reported large numbers of bacterial species in milk with no evidence of any inflammatory response. Similarly, the presence of bacterial pathogens commonly associated with intramammary infections in dairy cattle (*Staphylococcus aureus* and *Streptococcus uberis*) in healthy samples suggests they could be part of the normal milk microflora (Oikonomou *et al.*, 2014).

In summary, published studies to date on the composition of the milk microbiome suggest the presence of a diverse commensal microflora. Both similarities and differences in community composition have been identified between normal and diseased milk, mammary gland quarters, over time and between different anatomical sites and host species.

1.2 Intramammary infections

Intramammary infections are the result of inflammation of the mammary gland. This is usually as a result of a bacterial infection. The term mastitis refers to when disease occurs as a result of infection (Harmon, 1994). Mastitis is often caused by one species of bacteria that might be part of a normal microbial community in the mammary gland.

1.2.1 Presentations of mastitis

The presentation of mastitis in sheep can be defined by severity, clinical signs or type of bacterial infection. When severity is used to define mastitis, subclinical and clinical are commonly used terms. When duration is used to describe mastitis, acute (rapid) and chronic (long lasting) are used.

Subclinical mastitis has no outward signs of disease and does not lead to any changes in the milk or udder appearance (Harmon, 1994). It is therefore identified through changes in milk composition, causative bacterial species in the milk or the somatic cell count (SCC) (Albenzio *et al.*, 2002; Keisler *et al.*, 1992; Watkins *et al.*, 1991). Subclinical mastitis can persist for long periods as it can remain undetected. Subclinical mastitis can also occur because clinical mastitis is not detected by farmers. The point at which clinical and subclinical mastitis are defined is therefore highly subjective.

Clinical mastitis is diagnosed visually through changes in the udder, milk or animal behaviour (Mavrogenis *et al.*, 1996). Clinical manifestations of mastitis are categorized into three grades in cattle (Green *et al.*, 1997). Type one is an acute local infection, type two is a systemic acute infection and type three a toxic infection. Coliforms have been associated with the type three toxic infection, although not all coliform infections lead to clinical mastitis, and up to 50% of such infections cause only mild forms of mastitis (Green *et al.*, 1998). These distinctions are also present in sheep, although depend on farmer observation to detect them.

1.2.2 Causative agents of mastitis

Many species of bacteria cause mastitis in sheep. Watts (1988) states that more than 130 organisms have been reported to infect the mammary gland of dairy cattle and it is likely that this is a similar number for sheep. The range of bacterial species detected in sheep milk during microbiological and molecular analysis (Smith *et al.*, 2011) suggests that a mixed community of bacteria are present in the mammary gland.

The major pathogens in sheep are; *Staphylococcus aureus*, *Streptococcus dysgalactiae*, *Streptococcus agalactiae*, *Streptococcus uberis*, *Escherichia coli* and *Mannheimia haemolytica* (Green *et al.*, 2005; Heras *et al.*, 2002; Mavrogenis *et al.*, 1996; Omaleki *et al.*, 2011; Rowe *et al.*, 2001; Watkins *et al.*, 1991). Coagulase-negative staphylococci (CNS) are considered the predominant minor pathogens because they are usually associated with milk and/or subclinical forms of mastitis. However, the ability of CNS to cause mastitis is thought to be variable depending on the pathogenicity of the individual bacterial species (Contreras *et al.*, 2007; Fthenakis and Jones, 1990; Pengov, 2001; Supré *et al.*, 2011).

1.2.3 Changes in microbial community structure linked to mastitis

Sometimes bacterial communities are only harmful when the balance of organisms changes. For example, foot rot in sheep (Calvo-Bado *et al.*, 2011) and gum disease in humans (Marsh, 1995). Bacterial species within a community may become harmful for a number of reasons. These include; a community change, the introduction of a new species, or an increase in abundance of a particular species. All of these could be detrimental to the host organism. One example of disease development as a result of a community change may be mastitis.

It has been proposed that subclinical and clinical mastitis are different stages in the progression of a single disease in sheep (Watson and Buswell, 1984). Watkins (1991) reported 38.5% of 26 mammary glands that developed clinical mastitis had a subclinical infection caused by the same bacterial species identified in the milk before the onset of clinical signs. This is also supported by research in dairy cattle where two-thirds of enterobacterial mastitis occurred in quarters infected with the same species of pathogen for up to 100 days (Bradley and Green, 2001a). In addition, high rates of detection of mastitis pathogens (~40% of dairy cattle tested) have been found in the dry period before a lactation (Green *et al.*, 2005).

These findings suggest that bacterial species persist in the mammary gland for long periods of time in subclinical states, possibly recrudescing to cause disease.

1.2.4 Associations between bacterial species in the mammary gland

Research conducted in dairy cattle has indicated positive correlations between the presence/absence of certain bacterial species and the subsequent risk of developing mastitis. For example, Green *et al.*, (2005) found the probability of isolating either *Escherichia coli* or *Streptococcus uberis* was significantly greater when the other organism was cultured in a milk sample. Green *et al.*, (2002) reported the presence of *Corynebacterium* species at drying

off in dairy cattle resulting in an increased risk of clinical mastitis, yet isolating the same species in the late dry period (after drying off) was associated with a reduction in the risk of clinical mastitis.

Such studies could suggest different bacterial species within a community have either synergistic or inhibitory influences on each other as previously suggested in relation to the anaerobes *Trueperella pyogenes* and *Fusobacterium necrophorum* (Pyorala *et al.,* 1992). Such findings also suggest that a community of multiple bacterial species could co-exist as a single unit and so changes within this community can result in disease.

1.3 Somatic cell count (SCC)

The SCC is the number of somatic cells, typically leukocytes, present per millilitre of milk (Lafi, 2006). The SCC is used as a method of monitoring milk health over time. It is used as an indicator of infection; a rise in SCC is indicative of a polymorphonuclear cell immune response to a bacterial intramammary infection (Mavrogenis *et al.*, 1996).

1.3.1 SCC in suckler sheep

The SCC has not been studied extensively in suckler sheep, so an accepted 'normal' value has not been determined. In dairy sheep, SCCs of 600,000 - 800,000 cells per ml are considered indicative of an infection, with up to 30% of new infections within a year associated with values in this range (Billion and Decremoux, 1998). However, this set level of SCC determines whether sheep milk can be sold for human consumption and there is no reason to believe that SCC would be this high. In recent research at the University of Warwick, SCC of sheep milk from a local farm was analysed and found to be similar in SCC to that of dairy cow milk. The SCC values for suckler sheep have been reported to be higher than that of milking sheep and cattle due to the lack of preventive management measures against subclinical mastitis (Gonzalez-Rodriguez. *et al.,* 1995).

1.3.2 SCC in dairy cattle

More established estimates are available for SCCs of dairy cattle due to the economic costs incurred as a result of reduced milk yield and quality as a result of infection (Deluyker *et al.,* 1993). In dairy cattle, Green *et al.,* (2006) reported that an SCC greater than 200,000 cells/ml significantly increased the risk of intramammary infection and SCCs below 100,000 cells/ml were indicative of a healthy mammary gland. Dohoo and Meek (1982) reported an SCC in the 100,000 - 300,000 cells/ml range to indicate an infection with minor pathogens with an

average SCC of 214,000 cells/ml as the threshold for uninfected quarters in cattle, both with and without a history of mastitis.

Low SCCs in dairy cattle herds (<150,000 cells/ml) have been attributed to a reduction in mastitis associated with contagious pathogens and improved control and management methods (O' Reilly *et al.*, 2006). However, published literature has also shown a high incidence of mastitis (36.7 quarter cases per 100 cow years) in dairy cattle herds with low bulk milk somatic cell count (BMSCC) (Peeler *et al.*, 2002). Cattle with a low SCC (<40,000 cells/ml) in the month before a case of clinical mastitis were shown to be more likely to have a severe infection compared to those with higher SCCs. This could suggest that cattle with higher SCC, which have been selected against, may have produced a stronger immune response to infection.

The presence of bacterial pathogens causing clinical mastitis with a reduced immune response from the host animal (low SCC), could suggest that these bacteria form part of a persistent community in the mammary gland. This could mean that the immune system is less equipped to recognise infection until it is in the clinical stages. Infection could also exist at such low numbers of bacteria in milk that the SCC and traditional culturing techniques fail to detect infection (Taponen *et al.*, 2009; Versalovic *et al.*, 1991). Some bacterial strains may have evolved resistance to treatment or the ability to evade the immune system and recrudesce to cause disease when there is a community change (Peeler *et al.*, 2002).

Research on the persistence of bacterial pathogens in the sheep mammary gland is very limited. Hence, the investigation of the bacterial genera present over time in relation to the SCC may provide insights into how disease develops, which could help determine novel approaches to disease management.

1.3.3 Increasing SCC with parity in dairy cattle

Parity (number of offspring; often used as a proxy for sheep age) has been associated with an increased risk of infection including severe clinical mastitis of the mammary gland in dairy cattle (Biffa *et al.*, 2005; Green *et al.*, 2002; Peeler *et al.*, 2002). Analysis of milk samples SCC suggested that SCC increases with age (Green *et al.*, 2005; Reneau, 1986). Conformational changes and/or damage to the mammary glands over time could increase the risk of disease (Green *et al.*, 2002). Alternatively, intramammary defence mechanisms may deteriorate with age (Green *et al.*, 2005). The increasing risk of disease with age could also indicate compositional changes in the community over time acting as a trigger for infection.

1.4 Methods to investigate bacteria in milk

1.4.1 Culture-independent methods

The approach chosen in this study to assess the composition and diversity of the sheep mammary gland is culture-independent. The gold standard to diagnose mastitis is the culture of bacteria from 2 samples of milk (Hogan *et al.*, 1999) although this is hotly disputed because a number of factors can affect the results of a bacteriological analysis. These include inappropriate media/culturing conditions, transient infections and intermittent shedding of organisms (Bishop *et al.*, 2010). Culture also lacks the discriminatory power to differentiate between closely related bacterial species (Zadoks and Watts, 2009). Such issues have resulted in conventional culturing failing to detect ~50% of clinical mastitis cases. Recently, culture-independent processes have identified mastitis pathogens in the same milk samples, often in substantial quantities (Kuehn *et al.*, 2013; Taponen *et al.*, 2009).

Consequently, culture-independent processes provide a more rapid and reproducible way to view complex microbial ecosystems with less bias of detection of one bacterial species over another. The culture-independent approaches used in this study have been used previously to profile communities (Braem *et al.*, 2012; Ercolini, 2004; Hunt *et al.*, 2011; Oikonomou *et al.*, 2014).

1.4.1.1 DNA extraction

In a culture-independent approach, DNA extraction is the first step in obtaining a representation of community composition. Therefore, the method used requires careful consideration to ensure representative DNA is obtained for subsequent analysis (Quigley *et al.*, 2012). The standard method for DNA extraction is alkaline phosphate separation of nucleic acids (Marmur, 1961). However, critical steps in which DNA extraction protocols can vary from this include; lysis, separation of nucleic acid from cell debris and sample substrate and purification method as detailed in Chapter 2.

1.4.1.2 Polymerase Chain Reaction (PCR) and Denaturing Gradient Gel Electrophoresis (DGGE) approach

Following DNA extraction, PCR is commonly used to amplify bacterial DNA. Despite some uncertainties in the phylogeny inferred from rRNA such as the rooting of different domains (Muyzer and Schäfer, 2001), the 16S rRNA approach remains the standard marker (Ludwig *et al.*, 1998) for bacterial identification using PCR. Bacterial 16S rRNA gene targeted PCR

products can then be used to visualise bacterial community diversity and composition using DGGE. DGGE is a molecular fingerprinting technique that separates unique nucleic acid species in polyacrylamide gels containing a linear gradient of DNA denaturants based on the electrophoretic mobility of partially melted double-stranded DNA (Muyzer, 1999).

PCR-DGGE has been used to characterise bacterial communities in milk (Braem *et al.*, 2012; Kuang *et al.*, 2009) and is often similar to results obtained using culture-dependent processes (Delgado *et al.*, 2008). However, there are some limitations. For example, bacterial species that constitute <1% of the total community cannot be visualised (Muyzer *et al.*, 1993). Only short sequences (up to ~500bp) can be visualised, limiting the amount of sequence information for identification in subsequent comparative analyses (Muyzer and Schäfer, 2001). Poor resolution of some DGGE bands can make band assignment and excision difficult and dissimilar sequences may co-migrate to the same position in a DGGE gradient, meaning 1 band may not correlate to 1 bacterial species (Muyzer and Schäfer, 2001; Rossello-Mora *et al.*, 1999). Similarly, multiple bands may represent the same bacterial species if some bacteria harbour more than 1 copy of the 16S rRNA encoding gene, meaning heterogeneous sequences are produced (Nubel *et al.*, 1996). Despite these limitations, DGGE offers a simple and reproducible approach for gaining an understanding of bacterial community composition and diversity.

1.4.1.3 High-throughput sequencing

High-throughput sequencing is revolutionising our understanding of the role microbial communities play in host health and disease (Proctor, 2011). Sequencing and analysis of hypervariable regions with the 16S rRNA gene can provide a relatively rapid and cost-effective method for assessing bacterial community diversity and composition and hence offers a methodology for the exploration of disease development (Oikonomou *et al.*, 2012).

There are several bench-top sequencing platforms currently being used. These include the 454 GS Junior (Roche), MiSeq (Illumina) and Ion Torrent PGM (Life Technologies). Published studies to date of the milk microbiota of dairy cattle have used barcoded-454 pyrosequencing (Kuehn *et al.*, 2013; Oikonomou *et al.*, 2014; Oikonomou *et al.*, 2012). All these platforms produce comparable data (Loman *et al.*, 2012). However, the MiSeq produces the lowest error rate, highest throughput per run, equivalent read quality and substantially reduced costs (Junemann *et al.*, 2013; Quail *et al.*, 2012).

Animal-based studies to date have only used high-throughput sequencing to analyse milk microbiota of dairy cattle, none have been used to investigate sheep milk.

1.4.1.4 Statistical modelling

Statistical models allow the investigation of multiple different variables in the form of mathematical equations. Modelling brings together components of a data set to identify associations and interactions between variables with adjustment for correlation between predictor variables. Statistical models have been used to investigate the pathogenesis of foot rot and the role of specific bacterial species (Witcomb *et al.*, 2014) and to elucidate interactions between bacterial species associated with intramammary infections (Green *et al.*, 2005).

1.4.1.5 Longitudinal studies

Longitudinal studies are epidemiological studies where the outcome of interest is measured repeatedly over time. As a result, they are time-consuming and complex to analyse because repeated observations over time are not independent (Twisk, 2003). However, they provide a powerful study design to provide evidence for causal associations between infection and disease. Interactions and behavioural patterns of causative agents of disease can be elucidated to improve understanding of disease initiation and progression.

1.5 Summary and conclusions from current knowledge

Investigating the composition and diversity of microbial communities is essential in improving our understanding of host health and disease. Research using culture-independent methods to date has identified a milk microbiota in both humans and dairy cattle that is not merely the result of skin contamination. Diverse and complex communities with both unique and shared organisms have been identified, as have differences between milk samples from animals in healthy and diseased states. Similarities and differences in the milk microbiota over lactation, between quarters and between animals as well as correlations between the presence/absence of certain bacterial species and microbiota across different anatomical sites have been identified.

However, no studies to date have been done on the mammary gland microbiome of suckler sheep. Published studies generally have not used many samples over time in a longitudinal data study which has the potential to identify bacterial species correlated to commensal,

protection, risk of infection and disease, to elucidate the complex, dynamic microbiology of the mammary gland.

Hence, a longitudinal study has been used to test the following hypotheses:

1. A natural microbial community (microbiome) forms in the suckler sheep mammary gland (Chapters 4 and 5).

2. Perturbations in the community result in disease (change in SCC) (Chapters 4 and 5).

3. With increasing number of lactations, the number and species of bacteria colonising the mammary gland increases (Chapters 4 and 5).

4. Differences in microbial community composition occur between mammary gland halves (Chapters 4 and 5).

5. Colonisation of the mammary gland is inevitable (Chapters 4 and 5).

The following objective was also established:

1. To develop effective molecular-based whole community approaches that were culture-independent. These included; DNA extraction, Polymerase Chain Reaction (PCR), Denaturing Gradient Gel Electrophoresis (DGGE) and high-throughput sequencing (Chapters 2, 3, 4 and 5).

Chapter 2 : Selection and optimisation of a method to extract DNA from sheep milk

2.1 Introduction

The selection of an appropriate DNA extraction method is crucial to obtain representative samples for analysis with culture-independent molecular techniques (Quigley *et al.*, 2012). The standard method for DNA extraction is alkaline phosphate separation of nucleic acids (Marmur, 1961). The critical steps in which DNA extraction can vary from this include; lysis method, separation of nucleic acid from cell debris and sample substrate and purification method. Lysis can be enzymatic, chemical, physical or a combination of these (Coppola *et al.*, 2001; Extramiana *et al.*, 2002; Flórez and Mayo, 2006; Meiri-Bendek *et al.*, 2002; Moschetti *et al.*, 1997; Odumeru *et al.*, 2001; Tilsala-Timisjärvi and Alatossava, 2004). Separation of nucleic acid from cell debris and sample substrate can be by phenol, alkaline phosphate or a combination of phenol and chloroform (Daly *et al.*, 2002; Giannino *et al.*, 2009; Millar *et al.*, 2000). Hydroxyapaptite (HTP) can also be used to separate nucleic acids from protein (Purdy, 2005; Purdy, 1996). Methods for purification of nucleic acids are numerous and include silica, ethanol, polyethylene glycol, isopropanol and sodium acetate (Extramiana *et al.*, 2002; Millar *et al.*, 2000; Odumeru *et al.*, 2001; Purdy, 2005). Kit-based methods often combine enzymatic lysis with silica-based nucleic acid purification and ethanol precipitation (Kuang *et al.*, 2009; Oikonomou *et al.*, 2012).

Sheep milk is a difficult substrate to process as it contains more fat and proteins than both cow and human milk which are inhibitory to PCR (Park *et al.*, 2007b; Rossen *et al.*, 1992). DNA extraction methods can also utilise compounds that can be PCR inhibitors such as detergents and alcohols (Rossen *et al.*, 1992). Milk samples used in this study are from sheep with no recorded signs of clinical disease and variable SCCs, which provides the additional issue of potentially low bacterial abundance as well as sample-to-sample variation. Hence, a method that could extract DNA from milk to process with PCR was essential (Bhatt *et al.*, 2012; Chui *et al.*, 2004; Cremonesi *et al.*, 2006; Luo *et al.*, 2007; Psifidi *et al.*, 2010; Ramesh *et al.*, 2002) in addition to practical considerations such as ease of protocol use, time to complete protocol and maximum number of samples that could be processed at one time (Riffon *et al.*, 2001).

In a previous study (Monaghan, 2010), five DNA extraction from milk methods were evaluated (Cremonesi *et al.*, 2006; Odumeru *et al.*, 2001; Riffon *et al.*, 2001; Tola *et al.*, 1997). Three methods used a combination of chemical (detergents and chaotropic agents) and physical (bead beating) lysis (Cremonesi *et al.*, 2006; Odumeru *et al.*, 2001; Tola *et al.*, 1997). A fourth method was based on centrifugation (Riffon *et al.*, 2001). A fifth method was kit-based and used an enzymatic (lysozyme and lysostaphin) based lysis followed by a silica column-based purification.

Each method was performed according to the published guidance and evaluated using a combination of nucleic acid purity and yield measurements determined by Nanodrop spectroscopy and bacterial 16S rRNA gene targeted PCR. Cremonesi *et al.,* (2006) produced inconsistent results, with duplicate DNA extractions producing DNA yields ranging from 7-180ng/µl. Riffon *et al.*, (2001) had a poor DNA extraction efficiency, with only 25% of duplicate samples producing PCR amplifiable DNA. Odumeru *et al.,* (2001) produced samples with low 260/280 ratios of 0.8-1.2. A 260/280 ratio of >1.8 indicates pure DNA, so ratios in the 0.8-1.2 range suggest contamination with milk components such as proteins and fats. The kit-based approach produced no visible PCR product for 6 sheep milk samples. The Tola *et al*., (1997) method was therefore selected as it consistently produced DNA that could be amplified by PCR, with 260/280 ratios in the 1.8-2.0 region. However, contamination in DNA extraction negative controls used with the Tola *et al.*, (1997) method were later detected by PCR.

Therefore, the previous study showed the importance of testing multiple DNA extraction methods as other authors have found (Psifidi *et al.*, 2010; Tomaso *et al.*, 2010) and testing for contamination. The result for using a kit-based approach was inconclusive, although both human and animal studies have used kits to assess microbial diversity (Oikonomou *et al.*, 2014; Oikonomou *et al.*, 2012; Oikonomou *et al.*, 2013). However, kit-based approaches are often not optimised for all types of samples (Psifidi *et al.*, 2010; Quigley *et al.*, 2012).

In this chapter, the testing of four different DNA extraction methods, 3 kit-based and 1 bespoke, to determine the most suitable to use with sheep milk for PCR analysis, are presented and discussed.

2.2 Materials and methods

2.2.1 Production of spiked samples

Spiked positive control samples were produced by adding a measured volume of one of the following substances; pasteurised cow milk, cow bulk tank milk, sheep milk or sterile water, to thawed bacterial cells. The thawed bacterial cells were *Staphylococcus aureus*, *Staphylococcus hyicus* or *Escherichia coli* and these were grown overnight in Luria Bertani (LB) agar at 37^{o}C. A loopful of bacteria was then inoculated into LB broth and incubated shaking at 37^{o}C overnight. Approximately 1ml of this liquid culture was transferred to a sterile 1.5ml microcentrifuge tube and centrifuged at 13,000 rpm for six minutes. The supernatant was removed and bacterial cells snap-frozen using dry ice and stored at -20^{o}C. When required, bacterial cells were thawed and mixed with one of the aforementioned substances. The resulting mixture was vortexed briefly to re-suspend cells. The volume and substance used varied depending on the DNA extraction method and this is described in the relevant section.

2.2.2 Testing four DNA extraction methods

Four DNA extraction methods were tested and/or optimised. These included 3 kit-based methods and 1 bespoke method. The kits used included the Nucleospin Blood and Tissue Kits (Macherey-Nagel Germany) and the Milk Bacteria DNA isolation kit (Norgen BioTek Corp, Canada). The bespoke method was adapted from Purdy (2005). Each method was trialled on at least two separate occasions. Results were evaluated using a combination of nucleic acid purity and yield measurements determined by Nanodrop spectroscopy (Thermo Fisher Scientific, Loughborough, UK), which is frequently used for the evaluation of DNA purity (Dekio *et al.*, 2005; Psifidi *et al.*, 2010), and bacterial 16S rRNA gene targeted PCRs. Both DNA yield and purity are important in an objective assessment of a DNA extraction method (Luo *et al.*, 2007).

2.2.2.1 Nucleospin blood kit protocol

The protocol was performed using manufacturer's instructions. This kit uses an enzymatic (proteinase K) and chemical (Buffers B1, B3 and BW contain guanidine hydrochloride) based lysis method, followed by a silica column-based purification. The protocol is rapid and easy to perform.

2.2.2.2 Nucleospin blood and tissue kits combined protocol

In an adaptation of the manufacturer's instructions, a bead-beating physical lysis step was included prior to an enzymatic (lysozyme, lysostaphin and proteinase K) lysis. DNA was purified via ethanol precipitation in a silica column-based purification as described by the manufacturer.

2.2.2.3 Norgen milk bacterial DNA isolation kit protocol

The protocol was performed according to manufacturer's instructions. The kit uses an enzymatic lysis (proteinase K, lysozyme, lysostaphin) followed by a resin-based column purification using ethanol.

2.2.2.4 Nucleic acid recovery from complex environmental samples (Purdy 2005)

The protocol was performed as stated in Purdy (2005; 1996). Briefly, the protocol uses a physical (bead-beating), detergent (Sodium dodecyl sulfate) and chemical (phenol) based lysis followed by nucleic acid separation from cell debris and sample substrate and purification using hydroxyapaptite (HTP) and Sephadex columns respectively. DNA is precipitated using ethanol and resuspended in sterile water. Extracted DNA was aliquoted into 10µl lots and stored at -20^{o}C.

2.2.3 Polymerase chain reaction (PCR) protocol and conditions

The DNA extraction methods tested were assessed with bacterial 16S rRNA gene PCRs. All PCR reactions were carried out under standard conditions on an Eppendorf master cycler in either a 15µl or 50µl final reaction volume. The reaction mixtures are presented in Table 2-1. A total of 3 different primer sets were used to assess DNA extraction results, the details of which are presented in Table 2-2. All PCR products were visualised by ethidium bromide-stained agarose gel electrophoresis in a 1% agarose gel made up with 1X TAE and run in the same buffer at 80-100 volts (depending on gel size) for 40 minutes.

Table 2-1: Standard PCR assay master mix with 2µl DNA [50ng/µl]

Master mix component	Working concentration	15µl reaction	50µl reaction
Primer Forward	20µM	0.75	1.5
Primer Reverse	20µM	0.75	1.5
Buffer	10X	1.5	5
$MgCl_2$	50mM	0.75	1.5
dNTPS	2mM	1.5	3
DMSO	100%	0.75	1.5
Platinum Taq	5U/µl	0.1	0.2
Water	-	6.9	33.8
Total (µl)	-	13	48

Table 2-2: Primers used to assess DNA extraction methods

Primer pair	Target site	Sequence 5'-3'	Annealing temperature (°C)	Product size (bp)	Reference
Uni 1870F Uni 2308R	Universal rRNA	TGGAAGGTTAAGAGGAGTGG GCCTCCGTTACCTTTTAGGA	59	438	(Riffon *et al.*, 2001)
PRBA338F PRBA518R	V3 region	ACTCCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGG	55	236	(Ovreas *et al.*, 1997)
341F 907R	-	CCTACGGGAGGCAGCAG CCGTCAATTCMTTTGAGTTT	54	566	(Muyzer and Schäfer, 2001)
27F 338R	V1-V2 region	AGAGTTTGATCCTGGCTCAG TGCTGCCTCCCGTAGGAGT	55	300	(Hunt *et al.*, 2011)

2.3 Results of testing and optimising selected DNA extraction methods

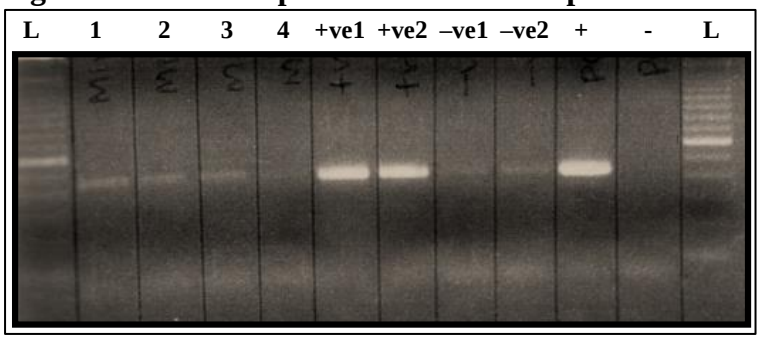
2.3.1 Nucleospin blood kit testing

A total of five trials were conducted using the Nucleospin blood kit with the protocol outlined in Section 2.2.2.1. Results from each of the five trials were assessed by general rRNA gene bacterial PCRs using primers Uni1870F/Uni2308R (Table 2-2). All trials tested 4 bulk tank cow milk samples (1-4), two spiked positive controls (*Staphylococcus aureus* and *Staphylococcus hyicus* in cow milk (+ve1 and +ve2)) and two negative extraction controls (water (-ve1) and phosphate buffered saline (PBS) (-ve2)) unless stated otherwise.

2.3.1.1 Trials 1-3 for the Nucleospin blood kit

Figure 2-1 illustrates the PCR results from Trial 1. Figure 2-1 shows faint but visible PCR amplification in cow bulk tank milk samples (Figure 2-1, lanes 1-4) and clear amplification in the two spiked positive controls (Figure 2-1, lanes +ve1 and +ve2). However, a faint band is also present in both DNA extraction negative controls (Figure 2-1, -ve1 and -ve2).

Figure 2-1: PCR amplification of Nucleospin blood kit trial 1 samples [1]



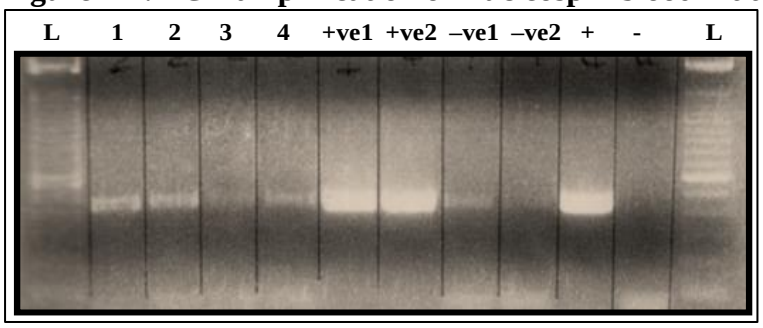
Due to the presence of a visible band in the DNA extraction negative controls for trial 1, all reagents and equipment were re-autoclaved and sterilised before trial 2 was completed. A physical lysis step (bead beating) of 0.5ml of cow bulk tank milk prior to starting the Nucleospin blood kit protocol was added. This step was included as physical lysis has previously been determined to increase DNA yield in other environmental samples (Leff *et al.*, 1998; Purdy, 2005).

Figure 2-2 shows the PCR results of trial 2. The greater intensity of each PCR product in comparison to those from trial 1 (Figure 2-1), indicated that trial 2 samples (Figure 2-2) had a higher PCR yield. Therefore, trial 2 results suggested that the physical lysis step increased the

[1]In Figure 2-1- Figure 2-5, PCR positive and negative controls are represented by '+' and '-' respectively. The DNA ladder is Hyperladder 1Kb (Bioline, UK) shown in each PCR image as 'L'.

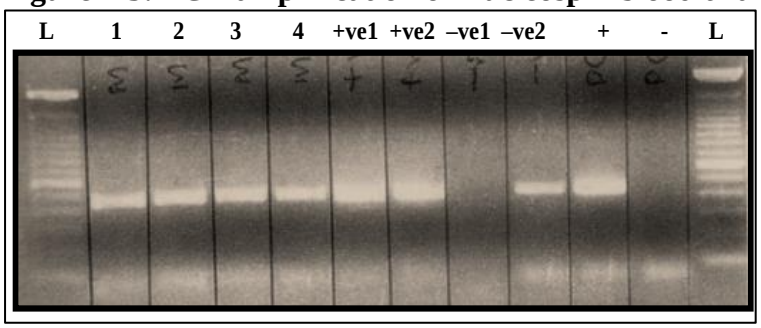
DNA yield. However, a faint band was still visible in one of the negative controls (Figure 2-2, -ve1), suggesting contamination remained.

Figure 2-2: PCR amplification of Nucleospin blood kit trial 2 samples



In trial 3, the volume of glass beads was halved to a quarter of the volume of a 0.2ml PCR tube. This was to investigate if reducing the amount of glass beads was feasible whilst retaining the effect of the physical lysis process. Figure 2-3 shows that in trial 3, contamination was again detected, but in the second negative control (Figure 2-3, -ve2) as opposed to the first in trial 2 (Figure 2-2, -ve1). This made it difficult to decipher the effect of reducing the volume of glass beads. However, the yield of the PCR reactions in trial 3 visually appears greater in comparison to trial 2, suggesting that the reduction in glass beads may have improved DNA extraction efficiency.

Figure 2-3: PCR amplification of Nucleospin blood trial 3 samples



2.3.1.2 Trials 4 and 5 of the Nucleospin blood kit

In trials 4 and 5, the starting volume of sample was increased to 300µl (Figure 2-4) and then 400µl (Figure 2-5). Increasing the starting volume to 300µl (Figure 2-4) was found to result in a decrease in PCR product yield (Figure 2-4, lanes 1-4) in comparison to trials 2 and 3 (Figure 2-2 and Figure 2-3, lanes 1-4). Trial 5 showed an increase in intensity of PCR product for cow bulk tank milk samples (Figure 2-5, lanes 1-4) in comparison to trial 4. The more intense contamination seen in trials 2 and 3 may explain why the PCR products in those trials

were more intense. However, the reduction in volume of glass beads used in trials 4 and 5 based on the results of trial 3, may have reduced the effectiveness of the physical lysis, with contamination in trial 3 masking the reduction.

Figure 2-4: PCR amplification of Nucleospin blood trial 4 samples

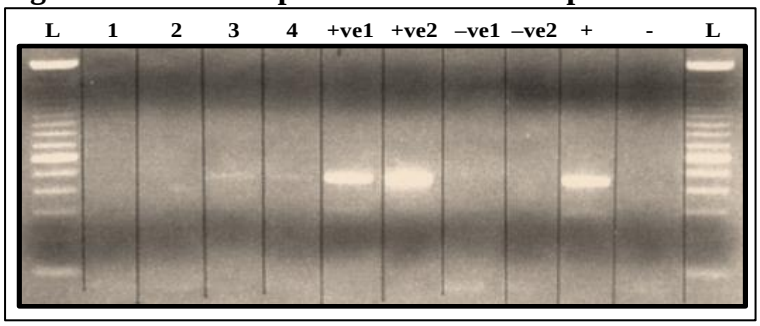
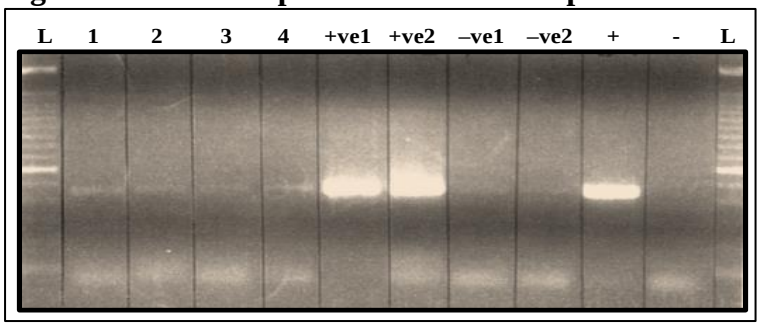


Figure 2-5: PCR amplification of Nucleospin blood trial 5 samples



2.3.2 Nucleospin blood and tissue kits combined protocol testing

Despite sterilisation and autoclaving of all reagents and equipment used in the Nucleospin blood protocol, the issue of contamination of DNA extraction negative controls remained. To try to improve the DNA extraction protocol by increasing yield and eliminating contamination, a new protocol was developed combining reagents and protocol steps from both Nucleospin kits. 10 trials were conducted on this protocol to test reproducibility and several modifications were made, of all which are described in the relevant section. All trial results were evaluated using Nanodrop Spectrometry and general bacterial 16S rRNA gene PCR.

2.3.2.1 Trials 1-4 of Nucleospin combined kits protocol tests

In trial 1 of the combined protocol, 3 bulk tank cow milk samples (1-3), 1 spiked positive control (*Staphylococcus aureus*) in pasteurised milk (+ve1) and 1 water only negative control (-ve1) were tested with spin columns from both the Nucleospin Blood and Nucleospin Tissue kits. Table 2-3 illustrates the Nanodrop results from trial 1.

Table 2-3: DNA concentrations [ng/µl], DNA yields and 260/280 ratios of samples from trial 1 of Nucleospin kit combined protocol

Sample identity	DNA concentration [ng/µl]	260/280 ratio	DNA Yield (µg)
Cow milk 1 (blood kit column)	6	1.3	0.6
Cow milk 2 (blood kit column)	5	1.2	0.5
Cow milk 3 (blood kit column)	5	1.5	0.5
Staphylococcus aureus spiked positive control (blood kit column)	31	1.8	3.1
Water only (blood kit column)	1	1	0.1
Cow milk 1 (tissue kit column)	2	1	0.2
Cow milk 2 (tissue kit column)	1	1.5	0.1
Cow milk 3 (tissue kit column)	1	1.4	0.1
Staphylococcus aureus spiked positive control (tissue kit column)	6	1.4	0.6
Water only (tissue kit column)	2	1	0.2

The 260/280 ratio is an indication of DNA purity as explained in Section 2.1. Table 2-3 indicates that higher DNA yields were achieved when the Nucleospin blood kit columns were used in the protocol in comparison to the tissue kit columns. For example, cow milk sample 1 processed with the blood kit column has a higher DNA concentration, 260/280 ratio and DNA yield in comparison to the same sample processed with a tissue kit column. This was corroborated by results seen when samples were visualised by agarose gel electrophoresis (not shown). Hence, trial 1 indicated that the protocol was most effective with the blood kit columns.

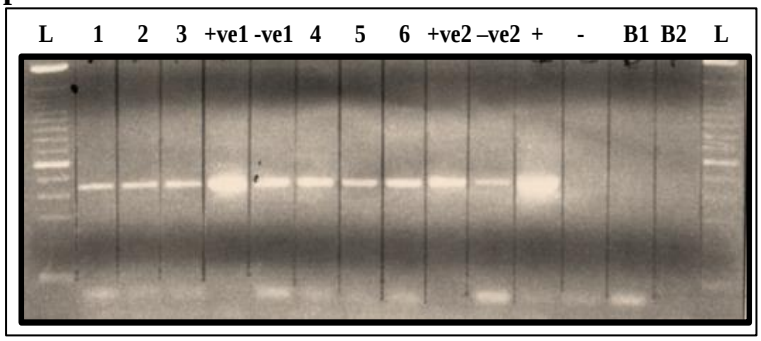
In trial 2, only the Nucleospin blood kit columns were used and a chloroform wash step was included after enzymatic lysis to see if this improved the removal of proteins from the milk. Table 2-4 shows that samples that included a chloroform wash had lower DNA concentration than those that did not e.g. cow milk sample 1.

When all samples from trial 2 underwent a general bacterial PCR as shown in Figure 2-6, all milk samples amplified to a similar extent (Figure 2-6, lanes 1-6). However, both DNA extraction negative controls still produced a positive PCR result (Figure 2-6, -ve1 and -ve2).

Table 2-4: DNA concentrations [ng/µl], DNA yields and 260/280 ratios of samples from trial 2 of Nucleospin kit combined protocol

Sample identity	DNA concentration [ng/µl]	260/280 ratio	DNA Yield (µg)
Cow milk 1 (chloroform wash)	7	1.6	0.7
Cow milk 2 (chloroform wash)	10	1.5	1
Cow milk 3 (chloroform wash)	9	1.6	0.9
Staphylococcus aureus in pasteurized milk (chloroform wash)	59	1.8	5.9
Water only (chloroform wash)	1	1.3	0.1
Cow milk 1 (no chloroform wash)	11	1.6	1.1
Cow milk 2 (no chloroform wash)	19	1.6	1.9
Cow milk 3 (no chloroform wash)	9	1.4	0.9
Staphylococcus aureus in pasteurized milk (no chloroform wash)	123	1.8	12.3
Water only (no chloroform wash)	0	-1	0

Figure 2-6: PCR amplification results for trial 2 of the combined Nucleospin kit protocol [2]

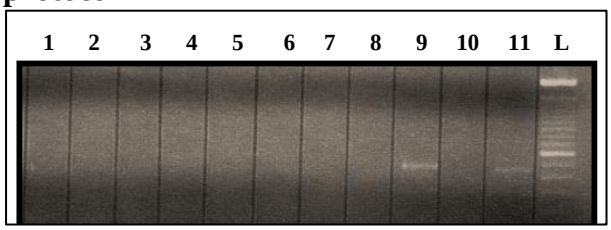


In trial 2, both the wash buffer and elution buffer were tested for contamination using PCR (Figure 2-6, B1 and B2 respectively) and found to produce no PCR product. In trial 3, a negative control was included for each step in the protocol to determine when contamination occurred. Post-DNA extraction, each negative control was tested using PCR as shown in Figure 2-7.

Figure 2-7 shows visible amplification in lanes 9 and 11 which refer to the steps after loading and washing each sample through the kit spin columns, suggesting the spin columns may be a contamination source.

[2] Numbers 1-6 in Figure 2-6 represent cow milk samples 1-3 with and without a chloroform wash respectively. DNA extraction positive and negative controls are labelled '+ve1', '+ve2' and '-ve1', '-ve2'. The PCR positive control is '+' and the negative control '-'. The protocol wash buffer (B1) and elution buffer (B2) were tested for contamination. Primers Uni1870F/2308R were used.

Figure 2-7: PCR results of negative controls from each step of the Nucleospin combined protocol [3]

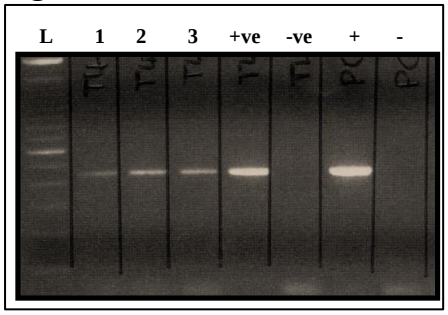


In trial 4, the Nucleospin blood columns and wash buffers were removed from the protocol and replaced with those from the DNeasy Blood and Tissue kit (Qiagen, Manchester, UK). The results of trial 4 are presented in Table 2-5 and Figure 2-8.

Table 2-5: DNA concentrations [ng/µl], DNA yields and 260/280 ratios of trial 4 of Nucleospin combined kit protocol using DNeasy kit spin columns

Sample identity	DNA concentration [ng/µl]	260/280 ratio	DNA Yield (µg)
Cow milk 1	5	1.4	0.5
Cow milk 2	11	1.4	1.1
Cow milk 3	2	1	0.2
Staphylococcus aureus in pasteurized milk (positive control)	23	1.6	2.3
Water only (negative control)	0	0	0

Figure 2-8: PCR results for trial 4 of Nucleospin combined kit protocol [4]



In comparison to samples from trial 2 (Table 2-4), the DNA concentrations and 260/280 ratios were lower in trial 4 which used the DNeasy kit spin columns. However, the PCR results in Figure 2-8 show visible PCR product in all cow milk samples (Figure 2-8, lanes 1-

[3] Numbers 1-11 refer to a step in the protocol after which a negative control was removed. Number 1 = post bead beating, 2 = post lysis buffer, 3 = post 37°C lysis incubation, 4 = post addition of proteinase K, 5 = post 56°C lysis incubation, 6 = post Buffer B3, 7 = post 70°C incubation, 8 = post addition of ethanol, 9 = post application to spin column, 10 = post Buffer BW, 11 = post Buffer B5. 'L' refers to Hyperladder 1kb (Bioline, UK).

[4] Numbers 1-3 are cow milk samples 1-3, '+ve' and '-ve' are the DNA extraction positive and negative controls and '+' and '-' are the PCR positive and negative controls respectively.

3) and no amplification in the DNA extraction negative control (Figure 2-8, -ve). This suggested that changing spin columns had removed the contamination source.

2.3.2.2 Trials 5-10 of combined Nucleospin kits protocol

Trial 5 of the protocol was a repeat of trial 4 to ensure that the result was reproducible and consistent. However, a faint PCR product was present in the DNA extraction negative control for trial 5 (not shown). Consequently, all DNA extraction reagents were tested for contamination using PCR and none were found to produce a positive result. Trials 6-10 tested variables other than spin column and reagents to identify the contamination source. The variables tested and their effects are presented in Table 2-6.

Table 2-6: Variables tested to identify contamination in Nucleospin kits combined DNA extraction protocol

Trial No.	Variable tested	Result
6	Use of heat block instead of water bath	Contamination still detected using PCR
7	Extraction carried out in flow cabinet	Contamination still detected using PCR
8	Only negative controls tested	Contamination detected using PCR when samples applied to spin columns
9	PCR primers changed to PRBA338f/PRBA518r (Ovreas *et al.*, 1997) and 341f/907r (Muyzer and Schäfer, 2001)	Contamination still detected inconsistently
10	Distilled water used as negative control changed to DNA/RNA free water	Distilled water contaminated and DNA/RNA free water not contaminated via PCR

As contamination was still being detected after the variables in Table 2-6 had been tested, this method was not investigated any further.

2.3.3 Norgen Milk Bacteria DNA isolation kit testing

The milk bacteria DNA isolation kit was tested in two trials. In trial 1, 4 bulk tank cow milk samples (1-4), two spiked positive controls (+ve1 and +ve2) and two negative controls (-ve1 and -ve2) were tested. Results were analysed using Nanodrop spectrometry (Table 2-7) and a bacterial 16S rRNA gene PCR using primers PRBA338F/518R (Table 2-2) as shown in Figure 2-9.

Table 2-7: Results from trial 1 of the Norgen Milk Bacteria DNA isolation kit

Sample identity	DNA concentration [ng/µl]	260/280 ratio	DNA Yield (µg)
Cow milk 1	20	1.8	2
Cow milk 2	21	1.4	2.1
Cow milk 3	24	1.3	2.4
Cow milk 4	19	1.4	1.9
Staphylococcus aureus in pasteurized milk (positive control +ve1)	29	1.6	2.9
Staphylococcus aureus in pasteurized milk (positive control +ve2)	27	1.6	2.7
DNAse and RNAse free water (negative control)	24	1.3	2.4

Figure 2-9: PCR results for trial 1 of the Norgen Milk Bacteria DNA isolation kit [5]

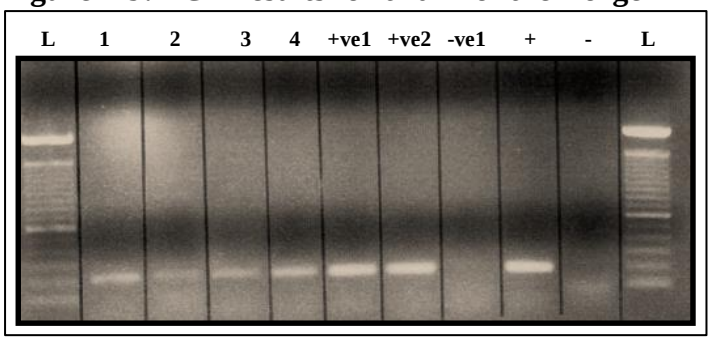


Table 2-7 shows a consistent DNA yield for the four bulk tank cow milk samples. However, the 260/280 ratios of 3 of the milk samples are in the 1.3-1.4 region which is low considering the DNA is being used for PCR. Despite this, Figure 2-9 shows amplification of DNA in all milk samples (Figure 2-9, lanes 1-4) and no amplification in the DNA extraction negative control (Figure 2-9, -ve1). However, when the PCR in Figure 2-9 was repeated, the DNA extraction negative control produced a faint but visible band which remained when the PCR primers were changed to 341F/907R (Table 2-2).

A second trial of the Norgen kit resulted in DNA samples with lower concentrations of 12-16ng/µl (Table 2-8) in comparison to the first trial (Table 2-7). Contamination of the DNA extraction negative control was seen in PCRs with both PRBA338F/518R and 341F/907R (Table 2-2). Due to the contamination issues, inconsistent DNA yields and difficulties in processing samples post-DNA extraction, investigation of the Norgen kit was terminated.

[5] PCR positive and negative controls are shown as '+' and '-' respectively.

Table 2-8: Results from trial 2 of the Norgen Milk Bacteria DNA kit

Sample identity	DNA concentration [ng/µl]	260/280 ratio	DNA Yield (µg)
Cow milk 1	16	1.4	1.6
Cow milk 2	17	1.7	1.7
Cow milk 3	12	1.8	1.2
Cow milk 4	13	1.4	1.3
Staphylococcus aureus in pasteurized milk (positive control)	16	1.4	1.6
Escherichia coli in pasteurized milk (positive control)	45	1.7	4.5
DNAse and RNAse free water	13	1.4	1.3

2.3.4 Testing of the Purdy (2005) protocol for DNA extraction

A method to recover nucleic acids from complex environmental samples has been developed (Purdy, 2005). The method has previously been used to extract DNA from sediment (Purdy *et al.*, 1997) so 2 bulk tank cow milk samples were tested with the method to see if DNA could be recovered from milk. When 5µl of each sample was run directly on a 1.5% agarose gel post-DNA extraction, a band of DNA was visible (Figure 2-10, lanes 1 and 2). The DNA from both milk samples also produced visible PCR product using both primers PRBA338F/518R and 341F/907R (Table 2-2). The PCR results using primers PRBA338F/518R are shown in Figure 2-11.

Figure 2-10: Agarose gel of DNA from two cow milk samples processed with the Purdy (2005) method

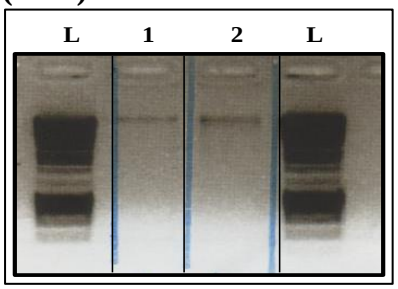
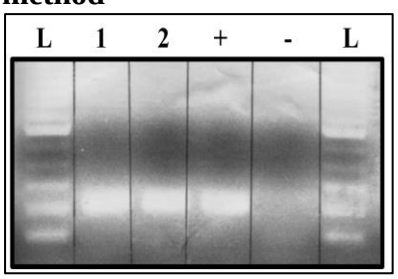


Figure 2-11: PCR products for two cow milk samples from trial 1 of the Purdy (2005) method



A second trial of the Purdy (2005) method was completed. Figure 2-12 shows the PCR results from trial 2. PCR product is visible in 3 of the 4 milk samples and the DNA extraction positive control (Figure 2-12, lanes 2-4 and +ve) with no contamination detected in the extraction negative control (Figure 2-12, -ve). No PCR product is visible in the first milk sample (Figure 2-12, lane 1) due to user error loading the agarose gel. Nanodrop spectrometry readings for both trial 1 and 2 are shown in Table 2-9.

Figure 2-12: PCR results for trial 2 of Purdy (2005) method

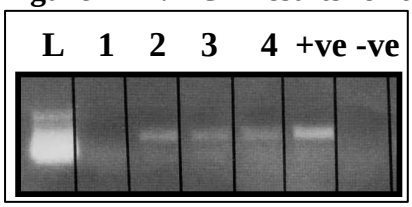


Table 2-9: DNA concentrations [ng/µl], DNA yields and 260/280 ratios for cow milk samples from trials 1 and 2 of the Purdy (2005) method [6]

Sample identity	DNA concentration [ng/µl]	260/280 ratio	DNA Yield (µg)
Cow milk 1	*44*	*1.5*	*4.4*
Cow milk 2	*44*	*1.6*	*4.4*
Cow milk 1	29	1.7	2.9
Cow milk 2	248	1.4	24.8
Cow milk 3	174	1.3	17.4
Cow milk 4	21	1.6	2.1
Staphylococcus aureus spiked positive control	554	1.5	55.4
Negative control	82	1.4	8.2

The DNA yields shown for samples processed using the Purdy (2005) method were higher than any of the other tested methods (Table 2-9), although there was some variation between milk samples aliquots e.g. 29ng/µl for cow milk 1 in trial 2 and 248ng/µl for cow milk 2 in trial 2. The negative control for trial 2 (Table 2-9) had a relatively low 260/280 ratio of 1.4 and a high DNA concentration of 82ng/µl. This could indicate cross-contamination of the negative control during the DNA extraction process, but this seems highly unlikely as the PCR of the negative control shown in Figure 2-12 is clear. The negative control was PBS. The phosphate in the PBS could have interfered with the absorbance readings at 260 and/280nm. This highlights the importance of using more than one method to visualise extracted DNA.

[6] Trial 1 milk samples are in italics.

Subsequent trials of the Purdy (2005) method showed consistent results, with DNA extracted repeatedly from milk samples with no contamination detected in negative controls, regardless of PCR conditions tested. Further PCR results from sheep milk samples processed using the Purdy (2005) method are presented in Chapter 4 and Appendix 2.

2.3.5 Sensitivity testing of the Purdy (2005) protocol

Sensitivity testing was carried out on the Purdy (2005) protocol to ensure that DNA could be extracted from milk samples with low bacterial abundance. Sensitivity testing used pasteurized milk spiked with either *Escherichia coli* K12 or *Staphylococcus hyicus*. Briefly, a ten-fold dilution series to 10^7 of the two cultures was made. Subsequently, 5µl drops of each dilution were spotted on LB agar plates and grown overnight at 37^{o}C, after which colonies were counted at a suitable dilution to determine the colony forming units per millilitre (cfu/ml) in the spiked pellet.

DNA was extracted from each dilution using the Purdy (2005) protocol. Each sample was then analysed using PCR. Sensitivity testing using the aforementioned protocol was completed a total of five times, with results presented in Table 2-10.

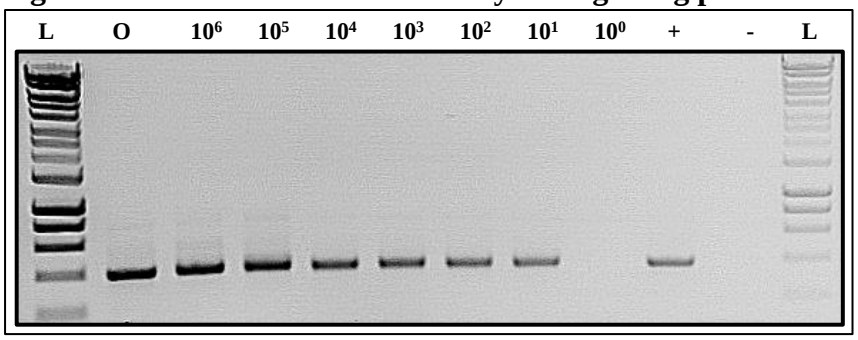
Table 2-10: Results of sensitivity testing of Purdy (2005) protocol

Test No.	Bacterial species	Average colony number in 5µl	Dilution factor to determine cfu/ml	Cfu/ml in spiked pellet	PRBA F/R PCR result (cfu/ml)	341F/907R PCR result (cfu/ml)	27F/ 338R PCR result (cfu/ml)
1	*E.coli* K12	1	10^4	2 x10^6	10^3	10^1	-
2	*E.coli* K12	3	10^5	6 x10^7	10^3	10^1	-
3	*E.coli* K12	4	10^5	4 x10^7	10^3	10^1	-
4	*S. hyicus*	2	10^5	2 x10^7	10^3	10^1	-
5	*E.coli* K12	10	10^4	2 x10^7	N/A	N/A	10^1

In the first 3 tests, *Escherichia coli K12* was used and results were analysed using primers PRBA338F/518R and 341F/907R (Table 2-2). PCR amplification was detected to 10^3 cfu/ml for PRBA338F/518R and 10^1cfu/ml for 341F/907R. The same result was produced when *Staphylococcus hyicus* was used. After extensive optimisation of the PCR protocol (detailed in Chapter 3), sensitivity testing was repeated a fifth time using PCR primers 27F/338R (Table 2-2) to ensure sensitivity remained consistent using the PCR protocol developed to

process all sheep milk samples. PCR product was detected to 10^1 cfu/ml (Figure 2-13), indicating a high level of sensitivity in the optimised PCR assay.

Figure 2-13: PCR result of sensitivity testing using primers 27F/338R [7]



[7] The original spiked milk sample is shown as 'O' and 10^6-10^0 refer to the cfu/ml of each of the samples in the dilution series. The PCR positive control is '+' and negative control '-' and 'L' is Hyperladder 1kb (Bioline, UK).

2.4 Discussion

2.4.1 DNA extraction method selected for further use

Of the four DNA extraction methods tested, the Purdy (2005) method was determined to be most suitable for use on milk samples for PCR analysis. The Purdy (2005) method was the only method tested where DNA was consistently extracted from milk, with the highest PCR and DNA yields and no contamination of DNA extraction negative controls. The results shown in this Chapter are in concordance with reproducibility studies previously carried out on this method (Purdy, 1996). The three kit-based methods tested had consistent contamination issues with made them non-viable for use in this study.

Cow bulk tank milk was used to investigate the DNA extraction methods due to the ease of its availability in comparison to sheep milk. Testing DNA extraction methods with cow milk also allowed sheep milk samples to be retained for DNA extraction with the method selected from initial tests.

2.4.2 Kit-based methods

Kit-based methods are convenient and easy to use and have previously been used in microbial diversity studies in both human and cow milk (Cabrera-Rubio *et al.*, 2012; Hunt *et al.*, 2011; Oikonomou *et al.*, 2014; Oikonomou *et al.*, 2012). A total of seventeen trials on 3 kit-based methods were detailed. Throughout these trials, several variables were altered to improve DNA extraction yield. These included; adjusting the starting volume of milk, adding physical lysis steps and changing spin-columns. Additional variables not detailed were also tested, including the incorporation of enzymatic lysis steps through the inclusion of proteinase k and alcalase. However, the effectiveness of many of these modifications were difficult to assess due to inconsistent contamination of negative controls detected using 16S rRNA bacterial PCRs. Attempts were made to address contamination by replacing and/or sterilising all reagents, although contamination was still detected, with only the kit spin columns identified as an irregular contamination source.

2.4.3 Contamination in kit-based methods reported in published literature

Contamination was detected in this study using PCR after a negative control was added to a spin column (Figure 2-7). Published literature from human studies has also pointed to DNA extraction columns in kit-based methods producing false-positive results from contamination with murine DNA (Hue *et al.*, 2010; Oakes *et al.*, 2010; Robinson *et al.*, 2010; Sato *et al.*,

2010). Furthermore, Erlwein *et al.,* (2011) investigated contamination with murine sequences when two different laboratory investigators in separate laboratories tested three different spin-column based kits. As found in this investigation, none of the buffers used in the kits were found to be contaminated when tested via PCR. However, a confounding issue in both circumstances is that often such buffers contain substances that inhibit PCR reactions, making such tests unreliable. Despite this, Erlwein *et al.,* (2011) soaked sections of spin column in elution buffer and found these produced a PCR signal, when elution buffer alone did not. This suggests that the columns themselves can be contaminated.

In addition, Naccache *et al.,* (2013) screened 8 different spin-column DNA extraction kits from 2 manufacturers for a highly divergent DNA virus identified via next generation sequencing which was thought to have a potential role in chronic seronegative hepatitis. Contamination of negative controls with the aforementioned virus occurred in kits produced by one manufacturer. Hence, the newly discovered virus was in fact a laboratory contaminant. It was hypothesized that as the virus is environmental and found in ocean water, the spin columns may have been contaminated during manufacture, which agrees with the opinion of Evans *et al.,* (2003) who reported contamination of Qiagen DNA extraction kits with *Legionella* DNA.

The high efficiency of silica-based spin columns in concentrating DNA during the extraction process (Boom *et al.,* 1990) may mean even trace contaminants are amplified. In relation to this, one study tested 4 different DNA extraction kits using a known pure DNA culture and identified via next-generation sequencing that the relative proportion of contaminant reads increased with each increasing dilution (Nick Loman, personal communication). Hence, this suggests an issue for low biomass samples, as DNA extracted using a contaminated kit-based method may result in a greater yield of contaminant than the actual organism(s) found in the sample.

In addition, Naccache *et al.,* (2013) and Tuke *et al.,* (2011) found contamination to be sporadic i.e. variable with time and/or batch dependent so the variation seen here via PCR detection could be explained by kit batch variability. Contamination via PCR reagents has also been reported (Knox *et al.,* 2011; Tuke *et al.,* 2011). The potential for contamination via commercial reagents has particular relevance for microbiome-based studies using next generation sequencing. As sequencing depth and capability increases, even trace amounts of contaminants could produce significant numbers of reads. This highlights the necessity of

testing laboratory reagents routinely for contamination and the use of DNA extraction negative controls as standard practice.

2.4.4 Advantages and limitations of Purdy (2005) method

Sporadic contamination in kit-based methods resulted in testing of potential methods moving away from this approach. The ability of the Purdy (2005) method to consistently extract DNA without contamination was the key motivation in carrying this method forward. The sensitivity of detection of 10^1 cfu/ml indicated that the method was capable of extracting DNA from milk samples with a low bacterial abundance which could be the case for some samples investigated in this study. Furthermore, the Purdy (2005) method involves production of columns in-house in sterile conditions, which could reduce the potential for contamination issues resulting from the use of commercial spin-columns.

However, the method does have some limitations. These are mainly linked to time and equipment required to prepare columns and then carry out the extraction protocol. The column design also means that only sixteen samples can be processed per run, limiting the high-throughput capacity of the method. The use of the toxic chemical phenol may limit the potential transferral of this method to a commercial setting. Despite the findings in this Chapter, easy to use kit-based methods have also been found to produce similar if not better results than phenol-based methods (Psifidi *et al.*, 2010) making them a desirable option for high-throughput studies, although kits have also been found to be less efficient when used to process samples with low levels of bacteria (Quigley *et al.*, 2012).

There are solutions to some of the Purdy (2005) method limitations. For example, during the optimisation process, it was found that columns can be pre-made and stored at 4^oC for up to one week prior to use without reducing extraction efficiency. This reduces time spent preparing columns between extractions. Also, there are alternative chemicals to test that could be used to replace phenol which may be more appropriate for transferral of the method to a commercial setting e.g. guanidine thiocyanate (Cremonesi *et al.*, 2006). The high-throughput capability of the protocol could also be improved by converting each stage into a 96-well format. Research into the potential for producing columns in a 96 well format from this project has already lead to promising preliminary results which are currently being tested for use in other research projects with larger sample numbers.

Furthermore, once processed, samples have to be aliquoted prior to storage and undergo minimal freeze-thaw cycles for further downstream processing. However, this is more than

likely linked to the sample DNA than the extraction protocol. Milk is a difficult substrate to work with as it contains substances such as fats and proteins (Park *et al.*, 2007a; Park *et al.*, 2007b) that can inhibit downstream applications such as PCR. Such milk components could also affect the DNA concentration [ng/µl] readings obtained via Nanodrop spectrometry which could explain the variation seen when testing duplicate samples.

2.4.5 Future potential for Purdy (2005) method and conclusions

For future studies, there is great potential for developing the DNA extraction method selected in this Chapter into a high-throughput protocol for processing samples on a large scale in both academic and commercial settings through the conversion of the three main stages (lysis, purification and precipitation) to a 96 well format.

Future studies must recognise the need for testing multiple methods with the use of negative controls to reduce the potential for contaminating organisms resulting in skewed study results, particularly if samples are to be analysed using highly sensitive next-generation sequencing technology. If repeating the testing detailed in this Chapter, it would be advisable to select multiple methods and conduct simultaneous testing to maximise the efficiency of selecting a DNA extraction protocol. In addition, using multiple methods to assess the DNA extraction efficiency is necessary to provide a comprehensive evaluation of each method tested (Psifidi *et al.*, 2010). Furthermore, although this was not achieved here, sequencing contaminated negative controls could produce a rapid inference on the contamination source, again reducing time spent on the selection and optimisation process, whilst still carrying out a thorough assessment of potential protocols.

This Chapter highlights the importance of thorough testing of DNA extraction methods to ensure representative results. The efficiency of different DNA extraction methods can vary with sample-type, so it is important to select methods based on the components in the sample-type that could reduce DNA extraction efficiency. Use of negative controls throughout also proved essential in determining the most effective method. The subsequent chapters detail the processing of sheep milk samples post-DNA extraction with the Purdy (2005) method.

Chapter 3 : Optimisation of PCR-DGGE protocol to analyse the total bacterial community in sheep milk

3.1 Introduction

The molecular fingerprinting PCR-DGGE approach was chosen to provide evidence of a bacterial community in the sheep mammary gland.

Following DNA extraction, PCR is commonly used to amplify the region of interest (Quigley *et al.*, 2011). The most common target for bacterial identification is the 16S ribosomal RNA gene (Flórez and Mayo, 2006) as it is highly conserved with variable domains. 16S rRNA-targeted PCR primers amplify portions of the 16S gene based on conserved regions whilst hypervariable regions allow for identification of different microorganisms (Quigley *et al.*, 2011).

PCR products can then be used to visualise the bacterial community using Denaturing Gradient Gel Electrophoresis (DGGE). DGGE is a molecular fingerprinting technique that provides a profile of community diversity based on the separation of unique nucleic acid species (Muyzer, 1999). DGGE separates DNA fragments of the same length but differing sequence based on the electrophoretic mobility of partially melted double-stranded DNA molecules in polyacrylamide gels containing a linear gradient of DNA denaturants (Giraffa and Neviani, 2001; Ogier *et al.*, 2002). It allows the visualisation of many samples from different time points, making it a powerful tool for monitoring community dynamics (Muyzer *et al.*, 1993).

PCR-DGGE has previously been used to characterise the bacterial microbiota in cow milk effectively (Braem *et al.*, 2012; Kuang *et al.*, 2009; Quigley *et al.*, 2011; Raats *et al.*, 2011). It is often the preferred method due to the inherent bias from culture-dependent methods (Koskinen *et al.*, 2010). Furthermore, no bacterial growth has been detected in 30-50% of clinical and up to 40% of subclinical cases of mastitis in dairy cows (Bradley *et al.*, 2007; Kuang *et al.*, 2009) and PCR identification of bacterial pathogens in comparison to culture has a reportedly higher success rate (Koskinen *et al.*, 2010). Hence, the lack of knowledge on

the conditions in which many bacteria grow in their natural environments means that molecular methods provide a more accurate representation of the bacterial community (Ercolini, 2004).

PCR-DGGE is also often in agreement with results obtained using culture-dependent processes in addition to identifying uncultivable organisms (Delbes *et al.*, 2007; Delgado *et al.*, 2008; Delgado *et al.*, 2013; Ogier *et al.*, 2002). DGGE can distinguish between bacterial species, with the recovery of DNA sequence information via band excision. For example, Chen and Hwang (2008) were able to differentiate between four species of *Staphylococcus* in cow milk. Hence DGGE is a reliable, inexpensive and more sensitive method to assess community diversity in comparison to culture-dependent methods (Muyzer and Schäfer, 2001).

However, PCR-DGGE is not without its limitations. Only bacterial populations that make up more than 1% of the total community can be detected by PCR-DGGE (Muyzer *et al.*, 1993). Co-migration of fragments amplified from different bacterial species can occur (Kuang *et al.*, 2009; Muyzer and Schäfer, 2001) as well as the formation of uncharacterised artefacts that can complicate the identification of specific bands (Delgado *et al.*, 2013). Distinct bands may not always correspond to different bacterial species as Kuang *et al.*, (2009) found 3 bands that were all identified as *Lactobacillus lactis*. Some DGGE bands can be too faint to excise or fail to produce positive PCR amplification post-excision (Delgado *et al.*, 2008).

Other complications include variation in bacterial community composition according to sample storage (Raats *et al.*, 2011; Smith *et al.*, 2011). Inconsistent bacterial cell wall lysis in DNA extraction and preferential or differential amplification of rRNA genes by PCR may occur (Delbes *et al.*, 2007; Muyzer and Smalla, 1998; Reysenbach *et al.*, 1992). Chimeric PCR products from too many PCR cycles may lead to a misleading evaluation of the bacterial species present (Giraffa and Neviani, 2001). Use of different target regions of 16S rRNA and/or different DGGE conditions could lead to different resolutions of separation. Heteroduplex molecules can also form and contribute to the difficulties in interpretation of banding patterns (Muyzer and Smalla, 1998). Hence, optimisation of the PCR-DGGE protocol is necessary to produce a reliable comparison between multiple samples.

In this Chapter, details of the testing and optimisation of 18 PCR primer sets, 3 PCR master mixes and the DGGE protocol are presented and discussed.

3.2 Materials and methods

3.2.1 Milk samples used in optimisation process

Milk samples from four different sheep (A7, A17, A32 and A48; milk samples were from a 2009 longitudinal study on one farm) were selected to test and optimise the PCR-DGGE protocol. Milk samples were collected from each udder half over 8 consecutive weeks and stored at -20^{o}C until processed. Parity refers to the number of pregnancies a ewe has had and so can be used as a measure of mammary gland activity and a proxy for age. Sheep A7 and A17 were parity 1, sheep A32 parity 2 and sheep A17 parity 3. Culture of milk samples at the external laboratory QMMS (Easton Hill, Wells, Somerset, UK) was done on brain heart infusion (BHI) agar supplemented with 5% sheep blood. Along with parity and culture data, somatic cell count data (SCC; a measure of immune response by the animal) was used to identify potential milk samples with a mixture of similarities and differences in bacterial community composition for PCR-DGGE protocol optimisation. The information available for each milk sample is presented in Table 3-1- Table 3-4.

Table 3-1: Information collected on sheep A7 milk samples [8]

Sample No.	Half	Week	SCC	Log SCC	Bacteria detected	Growth
13	L	1	240000	5.38	*Bacillus, Staph*	S,S
251		2	120000	5.08	*Bacillus,* maybe *Nocardia, Staph*	S,S,S
495		3	62000	4.79	G-ve	S
709		4	20000	4.30	G-ve, *Staph*	S,S
815		5	40000	4.60	*Staph*	S
1023		6	276000	5.44	*Staph*	S
1151		7	200000	5.30	*Bacillus*	S
1283		8	160000	5.20	NG	-
14	R	1	314000	5.50	*Staph*	S
252		2	120000	5.08	*Bacillus, Proteus, Staph*	H,H,H
496		3	74000	4.87	*Bacillus,* G-ve, *Staph*	S,S,S
710		4	40000	4.60	*Bacillus, Staph*	S,H
816		5	86000	4.93	*Bacillus,* G-ve	S,S
1024		6	132000	5.12	*Bacillus, Staph*	S,S
1152		7	80000	4.90	*Staph*	H
1284		8	138000	5.14	*Bacillus, fungal,* G-ve	S,S,S

[8] In Table 3-1 - Table 3-4: 'Half' refers to sheep mammary gland half and 'L' and 'R' within the 'Half' column specifies the left or right mammary gland respectively. *Staph* = *Staphylococcus* spp., *Coryne* = *Corynebacterium* spp., *Bacillus* = *Bacillus* spp., maybe *Nocardia* = maybe *Nocardia* spp., G-ve = Gram negative bacteria, fungal = fungal organism. S = small growth of bacteria, M = medium growth of bacteria, H = high growth of bacteria, NG = no growth of bacteria.

Table 3-2: Information collected on sheep A17 milk samples

Sample No.	Half	Week	SCC	Log SCC	Bacteria detected	Growth
31	L	1	276000	5.44	*Coryne, Staph*	S,S
323		2	624000	5.80	*Bacillus, Coryne, Staph*	M,M,M
583		3	460000	5.66	NG	-
753		4	272000	5.43	*Staph*	S
849		5	356000	5.55	*Bacillus, Staph*	S,S
1053		6	420000	5.62	*Bacillus, Staph*, G -ve	S,S,S
1195		7	932000	5.97	*Bacillus, Coryne, Staph*	S,S,S
1707		8	1270000	6.10	*Bacillus*	S
32	R	1	146000	5.16	*Bacillus, Coryne*	S,S
324		2	470000	5.67	*Bacillus,* coliform, *Coryne, Staph*	M,M,M,M
584		3	444000	5.65	*Bacillus*	S
754		4	364000	5.56	*Staph*	S
850		5	4000	3.60	*Bacillus,* G-ve, *Nocardia, Staph*	S,S,S,S
1054		6	336000	5.53	*Bacillus,* G-ve	S,S
1196		7	380000	5.58	*Staph*	S
1708		8	1098000	6.04	*Bacillus, Coryne*	S,S

Table 3-3: Information collected on sheep A32 milk samples

Sample No.	Half	Week	SCC	Log SCC	Bacteria detected	Growth
61	L	1	110000	5.04	NG	-
245		2	64000	4.81	*Bacillus, Staph*	S,S
481		3	116000	5.06	*Bacillus,* maybe *Nocardia, Staph*	S,S,S
685		4	316000	5.50	*Staph,* G-ve	S,S
801		5	124000	5.09	*Staph*	S
1029		6	444000	5.65	*Bacillus, Staph*	S,S
1141		7	214000	5.33	*Bacillus, Staph*	S,S
1261		8	548000	5.74	fungal	S
62	R	1	106000	5.03	*Bacillus, Coryne*	S,S
246		2	66000	4.83	*Bacillus*	S
482		3	146000	5.16	*Coryne*	S
686		4	432000	5.64	*Staph*	S
802		5	112000	5.05	*Bacillus, Staph*	S,S
1030		6	408000	5.61	*Bacillus, Staph*	S,S
1142		7	70000	4.85	*Bacillus, Staph*	S,S
1262		8	410000	5.61	*Staph*	M

Table 3-4: Information collected on sheep A48 milk samples

Sample No.	Half	Week	SCC	Log SCC	Bacteria detected	Growth
93	L	1	220000	5.34	*Bacillus, Coryne, Staph*	S,S,S
297		2	144000	5.16	*Bacillus, Staph*	S,S
577		3	76000	4.88	*Staph*	S
747		4	82000	4.91	*Bacillus, Staph*	S,S
857		5	72000	4.86	*Bacillus,* G-ve, *Staph*	S,S,S
1069		6	42000	4.62	*Bacillus, Staph*	S,S
1203		7	46000	4.66	*Bacillus, Coryne, Staph*	S,S,S
1717		8	914000	5.96	G-ve	S
94	R	1	202000	5.31	*Bacillus,* maybe *Nocardia, Staph*	M,S,S
298		2	111000	5.05	*Bacillus, Coliforms,* maybe *Nocardia, Staph*	S,S,S,S
578		3	46000	4.66	*Staph*	S
748		4	142000	5.15	G-ve, *Staph*	S,S
858		5	38000	4.58	*Bacillus, Staph*	S,S
1070		6	34000	4.53	*Staph*	S
1204		7	68000	4.83	*Bacillus, Coliforms, Staph*	S,S,S
1718		8	70000	4.85	*Bacillus,* G-ve	S,S

3.2.2 PCR protocols

3.2.2.1 PCR primers

Eighteen PCR primer sets were assessed for use in the PCR-DGGE approach as shown in Table 3-5. PCR conditions for each primer set are detailed in Table 3-6.

Table 3-5: Details of ten PCR primers sets tested for use in the PCR-DGGE protocol to assess microbial diversity in sheep milk samples

Primer pair	Target site	Sequence 5'-3'	Product size (bp)	Reference
341F 518R 907R	-	CCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGG CCGTCAATTCMTTTGAGTTT	 250 566	(Muyzer and Schäfer, 2001)
PRBA338F PRBA518R	V3 region	ACTCCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGG	236	(Ovreas *et al.*, 1997)
8F 1541R	-	AGAGTTTGATCCTGGCTCAG AAG GAG GTG ATC CAG CC	800	(Baker *et al.*, 2003; Reysenbach *et al.*, 1994)
Epsilon 1541R	-	GAGASTTTGATCMTGGCTCAG AAGGAGGTGATCCAGCC	600	(Embley, 1991)
968F 1330R 1346R 1385R 1401R	-	AACGCGAAGAACCTTAC TAGCGATTCCGACTTCA TAGCGATTCCGACTT CA CGGTGTGTACAAGACCC CGGTGTGTACAAGACCC	420 362 378 417 433	(Muyzer and Schäfer, 2001; Nubel *et al.*, 1996)
1055F 1392R	V9 region	ATGGCTGTCGTCAGCT ACGGGCGGTGTGTAC	323	(Ferris *et al.*, 1996)
pA pAdeg pH	-	AGAGTTTGATCCTGGCTCAG AGASTTTGATCMTGGCTCAG AAGGAGGTGATCCAGCCGCA	1500	(Edwards *et al.*, 1989)
357F 518R	V3 region	CCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGG	233	(Muyzer *et al.*, 1993)
7F 1400R 1541R	-	AGAGTTTGATCMTGGCTCA GGTGTGTACAAGGCCCGG AAGGAGGTGATCCAGCC	 1393 1534	(Embley, 1991)
27F 1492R 1522R 338R	V1-V2 region	AGAGTTTGATCCTGGCTCAG GGTTACCTTGTTACGACTT AAGGAGGTGATCCANCCRCA TGCTGCCTCCCGTAGGAGT	600 1500 300	(Hunt *et al.*, 2011; Suzuki and Giovannoni, 1996)

Table 3-6: PCR cycling conditions for all 16S rRNA bacterial PCR primers tested

Primers	PCR program	Reference
341F/518R/ 907R	94°C for 5 min, followed by 20 cycles of 94°C for 1 min, 65°C-55°C (touchdown 2°C per cycle), 72°C for 3 min, followed by 15 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 3 min, followed by final extension at 72° for 7 min.	(Muyzer *et al.*, 1993; Muyzer and Schäfer, 2001)
PRBA338F/ 518R	94°C for 5 min, followed by 30 cycles of 1 min at 94°C, 1 min at 55°C, 1 min at 72°C, followed by 5 min final extension at 72°C	(Muyzer *et al.*, 1993)
8F/Epsilon/ 1541R	96°C hot start, 96°C for 2 min followed by 10 cycles of 96°C for 1 min, 50°C for 30 sec, 72°C for 2 min 30 sec, followed by 20 cycles of 94°C for 30 sec, 50°C for 30 sec, 72°C for 2 min, final extension of 72°C for 5 min.	(Embley, 1991)
968F*/1330R/ 1346R/ 1385R/1401R	94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 63°C for 1 min, 72°C for 1 min, followed by final extension at 72°C for 5 min.	(Muyzer and Schäfer, 2001; Nubel *et al.*, 1996)
1055F/1392R	94°C for 5 min, followed by 11 cycles at 94°C for 1 min, 1 min at 53°C-43°C (touchdown –1°C per cycle), 72°C for 3 min, followed by 20 cycles of 94°C for 1 min, 43°C for 1 min, 72°C for 3 min, final extension of 72°C for 7 min.	(Ferris *et al.*, 1996; Muyzer and Schäfer, 2001)
pA/pAdeg/pH	94°C for 1 min, followed by 40 cycles at 94°C for 1 min, 62°C for 1 min, 72°C for 1.5 min, followed by final extension at 72°C for 7 min.	(Edwards *et al.*, 1989)
357F*/518R	94°C for 5 min, 65°C for 1min followed by 20 cycles of 65°C-55°C (touchdown 1°C every second cycle), 55°C for 5 min, final extension of 72°C for 3 min.	(Muyzer *et al.*, 1993)
7F/1400R/ 1541R	95°C hot start, 95°C for 1 min, followed by 30 cycles of 95°C for 40 secs, 55°C for 40 secs, 72°C for 2 min, followed by final extension at 72°C for 6 min.	(Embley, 1991)
27F/1492R/ 1522R	94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, followed by final extension at 72°C for 5 min.	(Suzuki and Giovannoni, 1996)
27F/338R*	94°C for 5 min, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min, followed by final extension at 72°C for 2min.	(Hunt *et al.*, 2011)

*GC clamp added: CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG G

3.2.2.2 PCR master mixes

Three PCR master mixes were used in the PCR optimisation process. These were; Promega GoTaq Green master mix (Southampton, UK), Invitrogen Platinum Taq (Paisley, UK) and Bioline MyTaq (London, UK). The default master mix used was Invitrogen unless specified otherwise in the relevant section.

3.2.2.3 PCR master mix components

All PCR reactions were carried out under standard conditions on an Eppendorf master cycler in a 50µl reaction volume. The reaction mixture used for both Invitrogen and Promega master mixes was previously presented in Chapter 3, Table 2-1. The reaction mixture for Bioline MyTaq is presented in Table 3-7.

Table 3-7: PCR assay mixture components with 1µl DNA [50ng/µl] for Bioline MyTaq

Master mixture component	Working concentration	Reaction volume (50µl)
Primer Forward	20µM	0.5
Primer Reverse	20µM	0.5
MyTaq master mix	N/A	25
DNA	50ng/µl	1
Water	-	23
Total (µl)	-	50

3.2.3 DGGE protocol

DGGE was performed as described in Muyzer and Schäfer (2001) using a DCode universal mutation detection system (BioRad, USA). Polyacrylamide gels (16cm x 16cm x 1mm) consisted of 6% (v/v) polyacrylamide (37:5:1, acrylamide/bisacrylamide) in 1X TAE buffer with a linear 20-80% denaturing gradient (100% denaturant contains 40% (v/v) formamide and 7M urea). Electrophoresis was performed at 60^{o}C for 18 hours at 100 volts. Each gel was loaded with PCR-amplified samples with a custom reference ladder consisting of 7 reference bands produced via excision of bands from a DGGE gel of sheep milk DNA samples using the protocol described by Muzyer and Schäfer (2001). The excised bands underwent PCR and purified amplicons were mixed in equal amounts. The gel was stained with SYBR Gold (Invitrogen) and visualised on a UV GelDoc (Geneflow, UK).

3.3 Results

3.3.1 PCR optimisation results

To achieve sufficient PCR amplification minus contamination for DGGE analysis, several options were investigated. Primer set 341FGC/907R was the first to be tested. Table 3-8 details the optimisations tested for 341FGC/907R to achieve a consistent PCR yield.

Table 3-8: 341F/907RGC optimisations [9]

PCR optimisation	Effect of optimisation
Temperature gradient from 65^{o}C-51^{o}C	PCR product faint and smeared
Touchdown PCR program (Table 3-6)	Faint PCR product
Two rounds of PCR	Amplification in all milk samples and first round PCR negative control
Reduced PCR cycle numbers in touchdown program	Only positive controls visibly amplified
Promega GoTaq and Fermentas PCR master mixes tested	Promega GoTaq amplified all milk samples and both DNA extraction and PCR negative controls. Fermentas only amplified the positive control

3.3.1.1 Using a nested PCR approach

The optimisations of primer set 314FGC/907R in Table 3-8 failed to achieve sufficient amplification of bacterial DNA from sheep milk sample sets A17 and A48. Subsequently, a nested PCR approach was tested, using different bacterial rRNA gene targeted primer sets in a first PCR and the 341FGC/907R primers in a second PCR.

Eight primer sets were tested in combination with 341FGC/907R. All primer sets were tested with two PCR master mixes; one containing BSA and the second DMSO. BSA and DMSO are PCR additives used to increase the efficiency of amplification of high GC content DNA sequences (Farell and Alexandre, 2012). Figure 3-1 shows the results of a nested PCR using primer set 27F/1522R and 341FGC/907R.

After the first (27F/1522R) PCR, only the spiked DNA extraction positive control and PCR positive control produced visible amplification (not shown). After the second PCR (Figure 3-1), all samples showed visible amplification including the first round PCR negative control (Figure 3-1, -1) and excluding the second round PCR negative control. The other 7 primer sets as well as primers 338FGC/518R were also tested, but produced the same result as in Figure 3-1.

[9] All optimisations in Table 3-8 were carried out using sheep A17 and sheep A48 milk samples.

Figure 3-1: Results from nested PCR using 27F/1522R and 341FGC/907R [10]

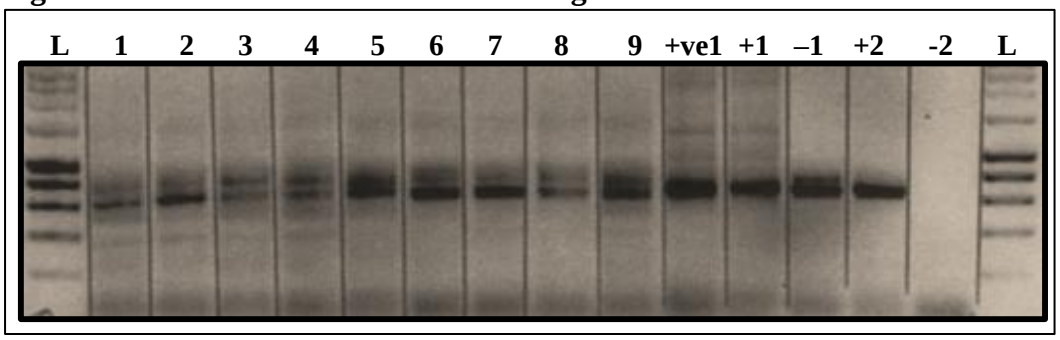


Table 3-9 summarises the remaining primer sets that were tested both individually and in a nested PCR approach with 341FGC/907R. None of the primer combinations in Table 3-9 produced sufficient PCR amplification without contamination. Hence, the primers listed in Table 3-9 and the nested PCR approaches were found to be unsuitable for the planned PCR-DGGE analyses.

Table 3-9: Nested PCR results for seven PCR primer sets tested with 341FGC/907R [11]

Primer set	Individual PCR result	Nested result
8F/1541R	Faint amplification	Only positive controls amplified
Epsilon/1541R	Only amplification in positive controls	Only positive controls amplified
27F/1492R	Only amplification in positive controls	Amplification in all samples apart from second round negative control
1055F/1392R	Only amplification in positive controls	Only positive controls amplified
7F/1400R/1541R	Only amplification in positive controls	Amplification in all samples apart from second round negative control
968FGC/1330R	Faint amplification	Amplification in all samples apart from second round negative control

3.3.1.2 Using a double round PCR approach

A further 7 primer sets were tested in single and double round PCRs to assess their ability to amplify DNA from sheep milk. The primer details and PCR results are summarised in Table 3-10. Overall, amplification was weak in one round of PCR and contamination occurred after

[10] In Figure 3-1, 'L' refers to Hyperladder 1kb (Bioline, UK). Numbers 1-9 are milk samples from sheep A17. The DNA extraction positive control is '+ve1' and the round 1 PCR positive control '+1'. The second round PCR positive control is '+2' and the second round PCR negative control is '-2'. The PCR additive used was BSA.
[11] All PCR primer sets were tested with sheep A17 and A48 milk sample sets.

a second round. This indicated that achieving sufficient amplification in one round of PCR was the most viable option to avoid false positives and contamination.

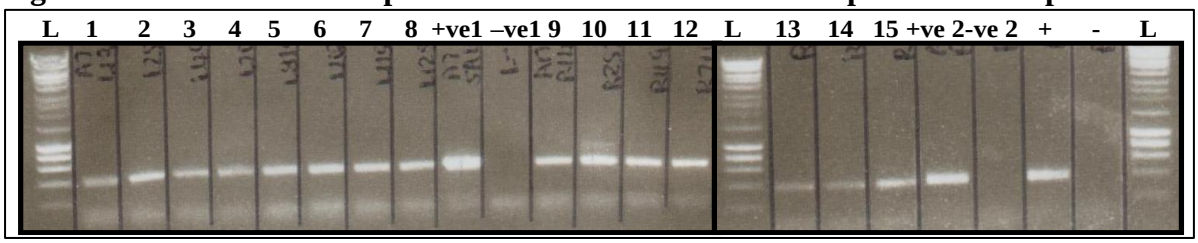
Table 3-10: Single and double round PCR results for five primer sets tested to amplify DNA from sheep milk

Primer set	Single PCR result	Double round PCR result
968FGC/1385R	No visible amplification in a single round of PCR for milk DNA tested	Amplification of milk DNA samples without false positives, but amplification still too weak for DGGE analysis
pA/pAdeg/pH	All samples amplified including DNA extraction negative controls	Same result as single round PCR, even when cycle number reduced and PCR master mix changed
968FGC/1346R/1401R	Only spiked DNA extraction positive controls with visible amplification when using 1346R. For 1401R, some faint amplification of milk DNA samples	Clear amplification from all milk DNA samples, with DNA extraction negative controls clear, but amplification of first round PCR negative control carried through to second round
338FGC/518R	Visible amplification in PCR positive control only	Amplification visible in majority of milk DNA samples, but also in DNA extraction and first round PCR negative controls
357FGC/518R	Faint amplification	Amplification of milk samples and PCR negative control from first round PCR

3.3.1.3 Primer set 27F/338RGC

Primers 27F/338RGC were tested in a single round of PCR with sheep A7 milk samples. As shown in Figure 3-2 (lanes 1-15), the yield of the PCR reactions was appropriate for DGGE analysis for all of the milk samples. To ensure this result was consistent, the PCR was repeated a further two times and the results remained the same.

Figure 3-2: PCR results for primer set 27F/338RGC with sheep A7 milk samples [12]



To see if PCR yield could be further enhanced in the 27F/338RGC PCR, two further optimisations were tested. In the first optimisation, the number of cycles in the PCR program was increased from 35 to 40. The second optimisation increased the amount of input DNA template from 1µl - 4µl. However, when compared to the amplification in Figure 3-2, PCR product yield appeared no greater and even lower in some milk samples for both optimisations. As a result, the published PCR conditions continued to be used (Table 3-6). To ensure the PCR yield seen with sheep A7 milk samples was consistent, primer set 27F/338RGC was also tested with sheep A32 milk samples as shown in Figure 3-3.

Figure 3-3: PCR results for primer set 27F/338RGC with sheep A32 milk samples [13]

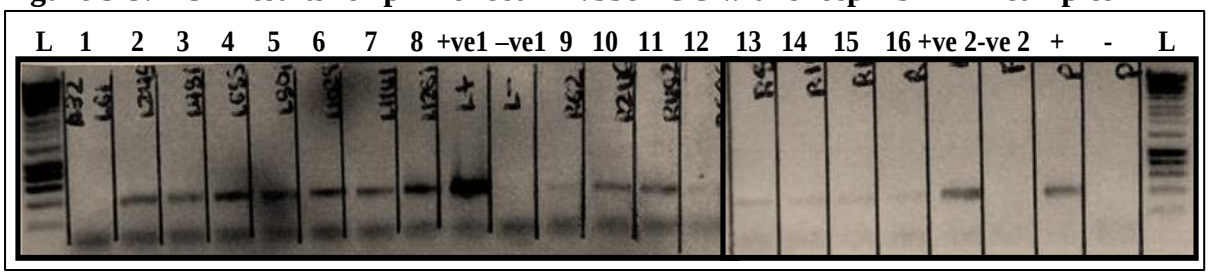


Figure 3-3 shows a good PCR yield for nine of the sheep A32 milk samples (Figure 3-3, lanes 2-8, 10-11) and weaker but visible PCR yield for 6 samples (Figure 3-3, lanes 9 and 12-16) and no amplification in 1 sample (Figure 3-3, lane 1). This result could indicate inconsistency in the PCR protocol, but seeing as this was the third time sheep A32 DNA samples had been freeze-thawed, there may have been an effect from sample handling. Sheep A7 DNA samples amplified more strongly in Figure 3-2 but had been aliquoted before storage at -20^{o}C meaning a new aliquot could be used for each PCR. There may also have been variation in bacterial load between sheep milk sample sets, resulting in differences in PCR amplification.

[12] In Figure 3-2, 'L' is Hyperladder 1kb (Bioline, UK). Numbers 1-15 are sheep A7 milk samples. The '+ve1' and '+ve2' are DNA extraction positive controls for 1-8 and 9-15 respectively. The '-ve1' and '-ve2' are DNA extraction negative controls for 1-8 and 9-15 respectively. The PCR positive control is '+' and the negative '-'.

[13] In Figure 3-3, 'L' is Hyperladder 1kb (Bioline, UK). Numbers 1-16 are sheep A32 milk samples. The '+ve1' and '+ve2' are DNA extraction positive controls for 1-8 and 9-16 respectively. The '-ve1' and '-ve2' are DNA extraction negative controls for 1-8 and 9-15 respectively. The PCR positive control is '+' and the negative '-'.

To date, PCR results with 27F/338RGC were consistent with no contamination or false positive issues. Overall, amplification appeared sufficient for subsequent DGGE analysis with only one round of PCR and multiple milk DNA samples from different sheep producing a positive PCR result.

3.3.1.4 Use of Bioline MyTaq with 27F/338RGC

Once testing was complete on primer set 27F/338RGC, a final step to improve product yield was a comparison of PCR amplification between the Invitrogen master mix used as standard and the Bioline MyTaq all-in-one master mix. Figure 3-4 shows the PCR products from a fifth sheep processed using an Invitrogen PCR master mix and Figure 3-5 is the same PCR using the Bioline MyTaq master mix.

Figure 3-5 shows a greater PCR product yield for all of the milk DNA samples in comparison to Figure 3-4. Five other milk sample sets were tested and compared in the same process and the Bioline MyTaq master mix had a consistently higher PCR product yield. Hence, the master mix of choice for use with primers 27F/338RGC was changed from Invitrogen to Bioline MyTaq.

Figure 3-4: 27F/338RGC PCR results using Invitrogen master mix [14]

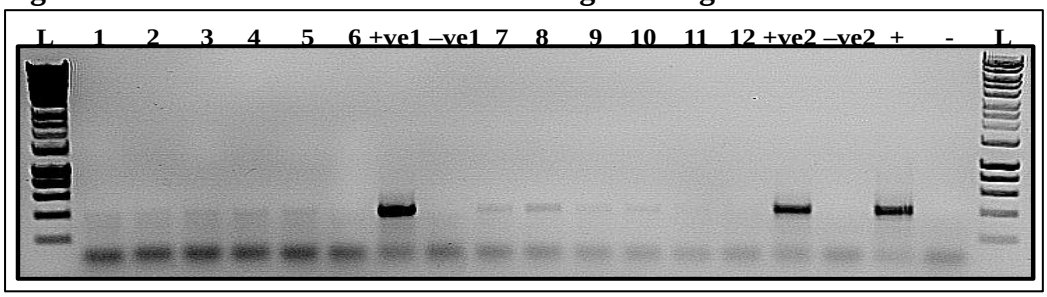
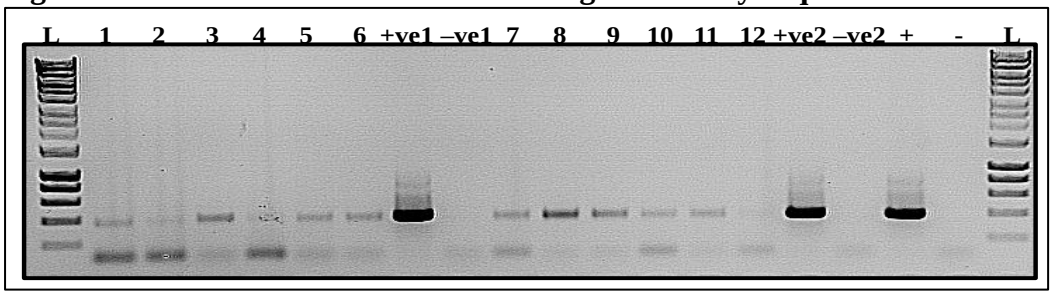


Figure 3-5: 27F/338RGC PCR results using Bioline MyTaq master mix



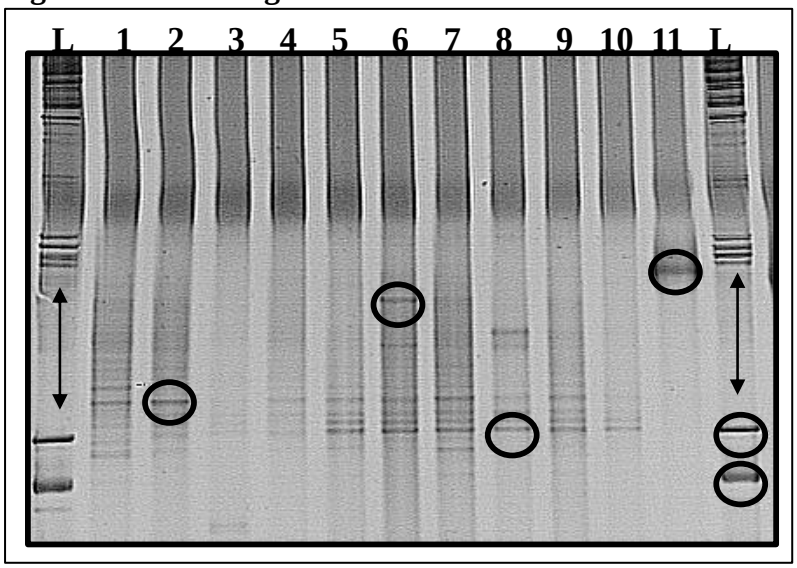
[14] In Figure 3-4 and Figure 3-5, 'L' refers to Hyperladder 1kb (Bioline, UK). Numbers 1-12 are milk samples from sheep A21 (detailed further in Chapter 4 and Appendix 1). The '+ve1' and '-ve1' are DNA extraction positive and negative controls for samples 1-6 and '+ve1' and '+ve2' are DNA extraction positive and negative controls for samples 7-12. The PCR positive control is '+' and the negative control '-'.

3.3.2 DGGE optimisation results

3.3.2.1 Custom DGGE reference ladder

Hyperladder 1kb (Bioline, UK) was initially used in DGGE gels as the reference ladder. It became apparent that several bands consistently appearing in PCR products from milk samples were present in a region of the reference ladder where there were no reference bands (as highlighted by arrows in Figure 3-6). Therefore, to accurately compare banding patterns across milk samples, a custom reference ladder was produced by excising a selection of DGGE bands from a test DGGE gel shown in Figure 3-6. The circled bands in Figure 3-6 were excised using the protocol detailed in Muyzer and Schäfer (2001). Excised bands underwent PCR and were run on a second DGGE gel to confirm the target band had been excised before purification, quantification and combining in equal concentrations to produce a reference ladder stock.

Figure 3-6: DGGE gel used to excise bands for custom reference ladder [15]



3.3.2.2 DGGE reproducibility

To ensure DGGE results were both representative and reproducible, 3 sets of milk samples underwent the PCR-DGGE process two consecutive times on different dates. Figure 3-7 and Figure 3-8 show 2 images, both of the same milk samples from sheep A20, processed separately using the same PCR-DGGE protocol. Overall, DGGE banding patterns were comparable for each sample with no significant differences, indicating that the PCR-DGGE protocol was sufficiently reproducible.

[15] In Figure 3-6, 'L' is Hyperladder 1kb (Bioline, UK), lanes 1-11 are PCR products from sheep A28 (details in Appendix 1).

Figure 3-7: DGGE image 1 of sheep A20 milk samples [16]

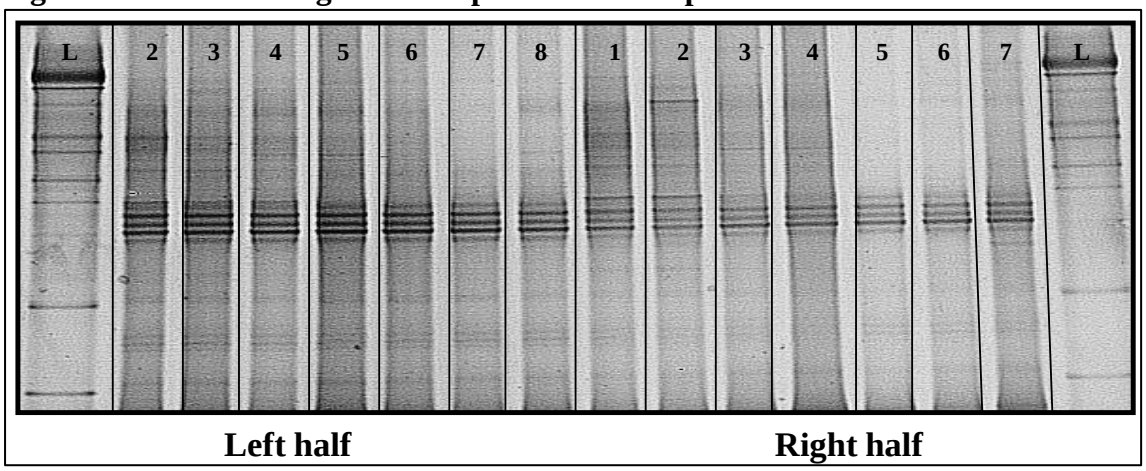
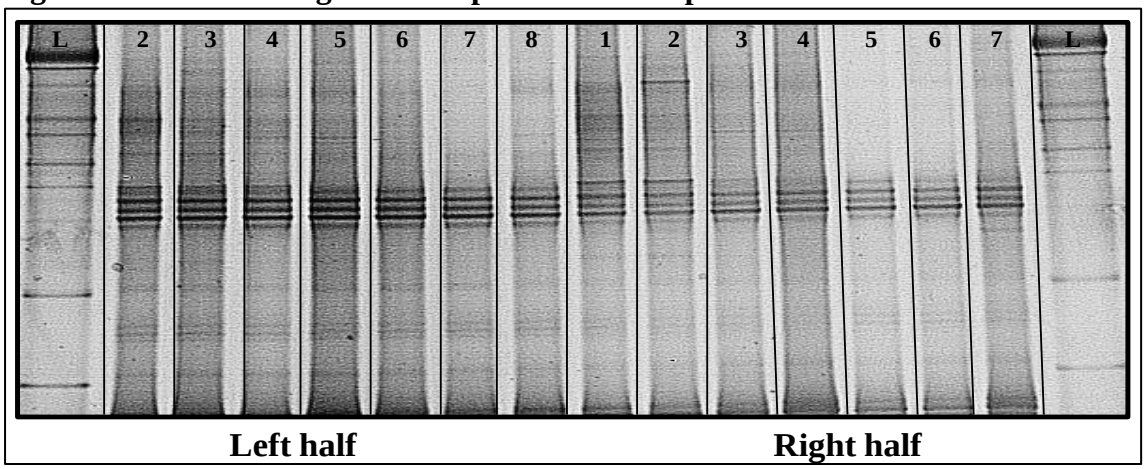


Figure 3-8: DGGE image 2 of sheep A20 milk samples



[16] In Figure 3-7 and Figure 3-8, 'L' is the custom reference ladder; numbers 1-8 refer to the week in lactation, with the mammary gland half defined in each Figure.

3.4 Discussion

3.4.1 PCR optimisation for DGGE

Optimisation of the PCR protocol was an extensive process as many variables were taken into consideration. All sheep milk samples used in this Chapter were from sheep with no recorded clinical signs of disease, so the bacterial load of some samples was potentially low. Reproducible PCR amplification can be difficult in low bacterial load samples as they may have insufficient template DNA for consistent PCR (Oros-Sichler *et al.*, 2006).

Kuang *et al.,* (2009) found achieving PCR amplification from non-clinical cow milk much more difficult in comparison to milk collected from mastitic cows, leading to the same conclusion that an overall lower bacterial load in milk makes processing more difficult.

The extensive optimisation required to achieve sufficient amplification of DNA supports this idea. The main issues in identifying an appropriate protocol included; poor PCR product yield, non-specific bands, large primer dimers and amplification of PCR negative controls; all of which can be related to the issue of processing low bacterial load samples which has been previously reported as challenging (Tanner *et al.*, 1998).

In relation to this, sample handling and storage also had to be carefully monitored because it became more difficult to produce PCR products from samples with successive freeze-thaw cycles as suggested in Section 3.3.1.3. Co-extracted substances from the DNA extraction protocol that could inhibit PCR could be reduced via DNA dilution or purification of PCR products, although this often resulted in too great a loss of DNA. By-eye visualisation of PCR products on a 1% agarose gel was therefore used to estimate the DNA concentration for loading onto DGGE gels.

To improve PCR product yield, several changes were made to PCR programs for different primer sets. These included changing the annealing temperature, reducing the number of cycles, doing a double round PCR and using a nested PCR approach. Reducing cycle number decreased the occurrence of contamination issues, but also resulted in decreased PCR product yield. Both the double round and nested approaches resulted in inconsistent results, with amplification of negative controls often occurring.

PCR product yield also varied with the primer set used and so each primer set had to be tested under standard conditions obtained from the literature before any potential alterations were

implemented. General bacterial primers that targeted the 16S rRNA gene had to undergo further optimisation of PCR conditions upon the addition of a GC clamp for DGGE analysis. Both Muyzer and Smalla (1998) and Kuang *et al.,* (2009) have reported that the addition of the GC clamp can reduce PCR efficiency.

The PCR reagents used also had an effect on PCR yield. Three PCR master mixes were tested; Promega, Invitrogen and Bioline. The Promega master mix tended to produce a larger PCR yield than Invitrogen, but was more prone to producing false positive amplification and amplification in DNA extraction and PCR negative controls. The Invitrogen mix was reliable in its ability to produce visible PCR product without false positives or contamination issues, but the PCR yield was often insufficient for subsequent DGGE analysis. The Bioline mix improved PCR yield in comparison to Invitrogen without the issues of false positive amplification and contamination, making it the master mix of choice.

The optimisation process highlighted the importance of using negative controls, for DNA extraction and PCR. Often samples would amplify a sufficient yield of PCR product but either both or a combination of the DNA extraction and PCR negative controls would also amplify. Hence, amplification in the milk DNA samples could not be relied upon to be representative of the bacterial community in the sample. Carrying negative controls through the PCR process therefore provided confidence in visible PCR amplification in samples representing what was present in the sample as opposed to background noise or contamination in the PCR. The extensive use of negative controls here provides a high level of confidence in the results obtained.

3.4.2 DGGE optimisation

Once the PCR protocol was optimised, it was necessary to ensure that both the PCR and DGGE protocols in combination would be suitable to produce images with banding patterns that could be analysed.

The DGGE gradient was trialled at both 20-80% and 30-80%. There was no significant difference in the distribution of banding patterns across the two gradients, although patterns were slightly more dispersed and therefore clearer with the 20-80% gradient so it was selected as the default.

The reference ladder has an important role in normalization of gels to make a comparison between gels feasible (Kuang *et al.*, 2009). When initially using Hyperladder 1kb with milk

DNA PCR products, several DGGE bands were present in a section of the ladder where there were no reference bands (highlighted by arrows in Figure 3-6). Consequently, subsequent comparison of bands within this area between samples was more complex and subjective. Hence a custom reference ladder was made by excision of a selection of bands from both hyperladder 1kb and milk DNA samples in a test DGGE to cover this region. Custom reference ladders have been produced for 16S rRNA amplicons previously (Braem *et al.*, 2012).

Determining an estimate of the DNA yield within each PCR product was necessary to optimise the volume of each sample to load onto the DGGE gel to achieve the standard of 300ng of DNA per lane (Muyzer and Schäfer, 2001). This facilitated a fair comparison between samples and also meant that subsequent data produced on the intensity of bands could be deemed reliable for analysis of the relative abundance of the same bacterial species in different sheep.

It was necessary to ensure that DGGE results were reproducible. Three milk sample sets underwent the PCR-DGGE process two consecutive times and produced comparable results. Kuang *et al.,* (2009) repeated PCR-DGGE analysis on five cow milk samples and the results were highly reproducible. In this study, 10% of all samples were run as technical replicates to check protocol reproducibility. Repeat DGGE images were comparable.

3.4.3 Chapter 3 conclusions

The work presented in Chapter 3 highlights the requirement to optimise methodology according to sample type. The use of bacterial DNA from sheep milk resulted in the testing of different PCR primers, conditions and reagents and the production of a custom DGGE reference ladder. Methods must also be reproducible, which is why PCR and DGGE methods used were tested on multiple milk samples from different sheep and were often repeated more than once to confirm results were consistent.

Chapter 4 : Longitudinal study of thirty sheep using PCR-DGGE

4.1 Introduction

Intramammary infections have a complex aetiology. In order to advance our understanding of these infections, it is necessary to monitor animals over time using longitudinal studies. Longitudinal studies have contributed to the understanding of infection and clinical disease. Bradley and Green (2001a) conducted a twelve month study into clinical mastitis in 6 dairy cattle herds and found *Escherichia coli* to persist within the mammary gland in excess of 100 days, with the same strains producing recurrent infections in different quarters. Green *et al.,* (2005) conducted a 14 month study where the prevalence of major bacterial species commonly associated with intramammary infections varied over time. Peeler *et al.,* (2002) collected data from 482 British dairy herds over 12 months to assess the overall incidence of clinical infection as well as the incidence in relation to parity and season.

If such studies used milk samples or specific data about an animal collected at only one time point, cause and effect could not be elucidated. Any significant effects detected may have occurred by chance, or due to animal, environmental or management strategies linked to that point in time, which could result in misleading conclusions. Hence, longitudinal studies allow changes and trends over time to be identified and help decipher the interactions and behavioural patterns of bacterial pathogens which can ultimately inform control programmes. It is likely that an intramammary infection and its clinical outcome are to some extent, determined by factors associated both with the animal and the bacteria, in addition to environmental and management factors (Green *et al.*, 2005), so longitudinal studies offer a method to investigate such multi-factorial infections.

Statistical modelling is a powerful tool in the investigation of bacterial species associated with disease. For example, Green *et al.,* (2005) used general linear mixed models of bacterial isolates from 480 dairy cattle milk samples to investigate the associations between the presence and absence of different bacterial species. The risk of isolating the major bacterial pathogen *Escherichia coli* was found to significantly increase when *Streptococcus uberis* was cultured from the same sample. In turn, the probability of isolating *Streptococcus uberis* from

a sample decreased significantly when *Corynebacterium* species were cultured, indicating that some bacterial species had either synergistic or inhibitory influences on other species. Such studies show how statistical modelling can be used to determine whether specific bacteria are linked to causing or protecting against intramammary infection to inform control strategies. Modelling can also be used to investigate disease initiation and progression. For example, Witcomb *et al.*, (2014) used a multinomial mixed regression model in a longitudinal study to investigate the roles of specific bacterial species in the infectious disease footrot.

In this Chapter, a longitudinal study of 30 sheep investigated using PCR-DGGE and statistical modelling is presented and discussed.

4.2 Materials and Methods

4.2.1 Longitudinal study samples

Milk samples were collected using standard techniques (Hogan *et al.*, 1999) from each mammary gland half of 30 sheep on one farm for up to 8 consecutive weeks in January and February 2010. The sheep were part of a flock of 220 mules on a commercial meat sheep farm in Wolverhampton, UK. Milk samples were placed on ice immediately and transported to the laboratory and frozen at -20^{o}C until used. Each sample was split before freezing into three aliquots. Two aliquots were submitted to the commercial laboratory QMMS (Easton Hill, Wells, Somerset, UK). One aliquot was for somatic cell count (SCC) analysis using the Fossomatic method (Delta CombiScope - Model FTIR 400, Drachten, Netherlands) and the second for microbiological culture. The third aliquot was retained for culture-independent molecular analysis.

The sheep in this study were part of a larger longitudinal study (Huntley *et al.*, 2012). They were free from clinical signs of mastitis during the sampling period. Milk samples from sheep that were part of the larger study were used during method optimisation to ensure protocols were sufficient for the specific sample type.

Table 4-1 summarizes the sheep identification, parity and number of milk samples collected per sheep. SCC data and results from bacteriological culture of each set of milk samples are in Appendix 1.

Table 4-1: Sheep identification and number of milk samples collected per sheep ordered by parity

Identification of sheep	Parity	Number of milk samples per sheep
A20	1	14
A35		11
A50		13
A15	2	14
A21		12
A26		13
A27		9
A28		10
A39		16
A4		10
A41		11
A44		13
A45		11
A46		15
A47		15
A5		14
A9		11
A12	3	13
A2		15
A22		14
A25		14
A29		14
A3		11
A16	4	14
A24		10
A40		13
A43		13
A49		12
A6	10	12
A23		12

In total, 379 milk samples from 30 sheep underwent PCR-DGGE analysis. Three sheep were parity 1, 14 were parity 2, 6 were parity 3, 5 were parity 4 and 2 parity ten.

4.2.2 DNA extraction

DNA was extracted from each milk sample using the method outlined by Purdy (2005) as previously detailed in Chapter 2 Section 2.2.2.4. DNA extractions were completed in batches of 8 or 16 samples with a *Staphylococcus aureus* spiked positive control and nuclease-free water negative control per batch.

4.2.3 PCR protocol for DGGE

Extracted DNA underwent PCR with primers 27F/338RGC that target the V1-V2 hypervariable segment of the 16S rRNA gene resulting in a 340bp amplicon suitable for phylogenetic classification of bacteria (Hunt *et al.*, 2011; Liu *et al.*, 2007; Muyzer and Schäfer, 2001). The primer sequences, GC clamp and PCR program were detailed previously in Chapter 3 Section 3.2.2.

All PCR reactions were prepared in a cabinet that was UV sterilised for 30 minutes prior to each use. A DNA extraction and no template PCR negative control were included in each PCR. All reactions were carried out using Bioline MyTaq under standard conditions on an Eppendorf master cycler. The PCR master mix components were detailed in Chapter 3, Table 3-7. Four 50µl PCR reactions were pooled per DNA sample prior to assessment using agarose gel electrophoresis. PCR products were assessed using 1% agarose gels stained with ethidium bromide and visualised using UV exposure on a Geldoc (Geneflow, UK). PCR products were then stored at 4^{o}C until further use.

4.2.4 Denaturing Gradient Gel Electrophoresis (DGGE)

DGGE analysis of amplicons was performed as described by Muyzer and Schäfer (2001) and previously detailed in Chapter 3 Section 3.2.3.

Pre-DGGE, DNA concentrations were estimated for each PCR product by comparison of PCR product intensity to known concentrations on a reference DNA ladder (Hyperladder 1kb, Bioline, UK) on a 1% agarose gel. The DNA concentration was then used to determine the volume of PCR product required to achieve 300ng/µl for each PCR amplicon as recommended by Muyzer and Schäfer (2001). PCR products were quantified so that samples could be compared to each other. The maximum volume of PCR product that could be added to a single lane of the DGGE gel was 55µl. Hence, for samples that required between 55-80µl for maximal loading, the maximum volume (55µl) was loaded. Samples that required more than 80µl were not loaded onto the DGGE gel, as any bands present would have been too weak to visualise.

All PCR amplicons for 1 sheep were loaded onto 1 DGGE gel to produce 30 DGGE images i.e. 1 image per sheep in the study. Selected DGGE bands were excised using the protocol in Muyzer and Schäfer (2001) for cloning followed by Sanger sequencing. All sequences were edited using the DNAstar SeqMan II sequence analysis package (Lasergene Inc). Band identity was determined using the National Centre for Biotechnology Innovation (NCBI)

standard nucleotide blast function with the 16S ribosomal RNA sequences (Bacteria and Archea) database (Altschul *et al.*, 1990).

4.2.5 Gel Compar II

DGGE gel images were edited for brightness and contrast using Adobe Photoshop CS6 (Adobe, UK) to reduce background noise that interfered with subsequent DGGE banding pattern recognition. DGGE banding patterns were analysed using the software package Gel Compar II version 5.1 (Applied Maths, Belgium) according to the instructions of the software provider. Briefly, each DGGE image was cropped and individual lanes defined and normalised. Normalisation minimized migration differences between gels by alignment of the reference ladder at either end of each DGGE gel. Optimal parameters for each DGGE lane were then adjusted to maximise the use of the automatic band recognition function to assign bands. DGGE profiles were then compared using a ranked Pearson-Product moment correlation coefficient and Unweighted-Pair Group Method with Arithmetic average (UPGMA).

4.2.6 Modelling of DGGE data

Data from the Gel Compar II analysis of DGGE images were imported into the software package MLwiN version 2.27 (Rasbash *et al.*, 2009). Data imported included sheep identification, parity, week of sampling, mammary gland half, milk sample identification and logSCC. The Gel Compar II analysis categorized each DGGE band into 1 of 35 band positions. Binary data on the presence or absence of a band for each milk sample and numerical data on the band intensity were analysed for the week of sampling and week prior to sampling in the model.

The model was a mixed effects regression model with logSCC as the dependent variable and sheep, mammary gland half and week as random variables. Parity was added as a fixed effect. A forward stepwise backward elimination approach with retention criteria $P < 0.05$ in the final model was used to determine the DGGE bands associated with a significant change in SCC.

The model took the form:

$$\text{Log(SCC)}_{ijk} = \beta_0 + \sum \beta x_{ijk} + \sum \beta x_{jk} + \sum \beta x_{k} + f_k + v_{jk} + u_{ijk}$$

where Log(SCC)_{ijk} = the log(SCC) on occasion ijk, where *i* is week, *j* is mammary gland half and *k* is sheep. β_0 = intercept, βx is a vector of fixed effects varying at level 1 (*ijk*), level 2 (*jk*), and level 3 (*k*), f_k, v_{jk} and u_{ijk} are the level 3, 2 and 1 residual variances, respectively.

4.3 Results

4.3.1 PCR results

PCR products were produced for all 379 milk samples in the longitudinal study using the protocol detailed in Chapter 3 Section 3.2.2. Figure 4-1, Figure 4-2 and Figure 4-3 are exemplars of the PCR results obtained for the 3 parity 1 sheep. The PCR results for all sheep are summarised in Table 4-2 and Table 4-3, with the PCR images in Appendix 2.

Figure 4-1: PCR products for parity 1 sheep A20 [17]

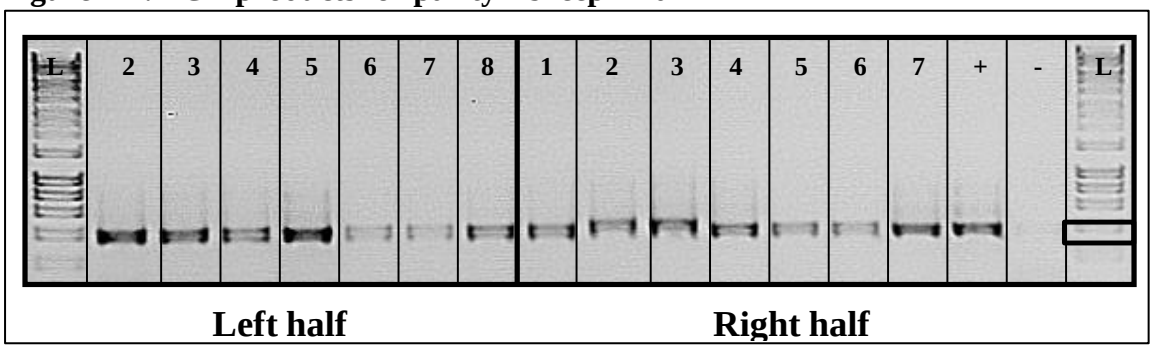


Figure 4-2: PCR products for parity 1 sheep A35

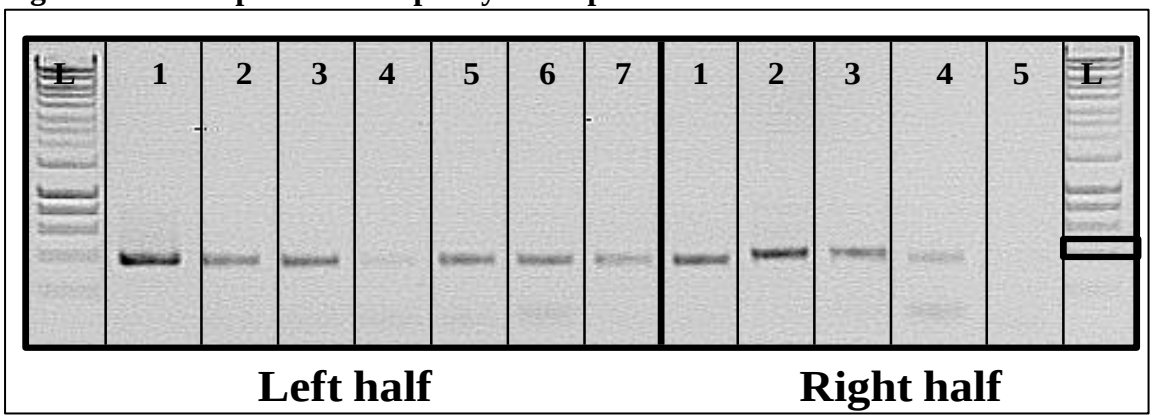
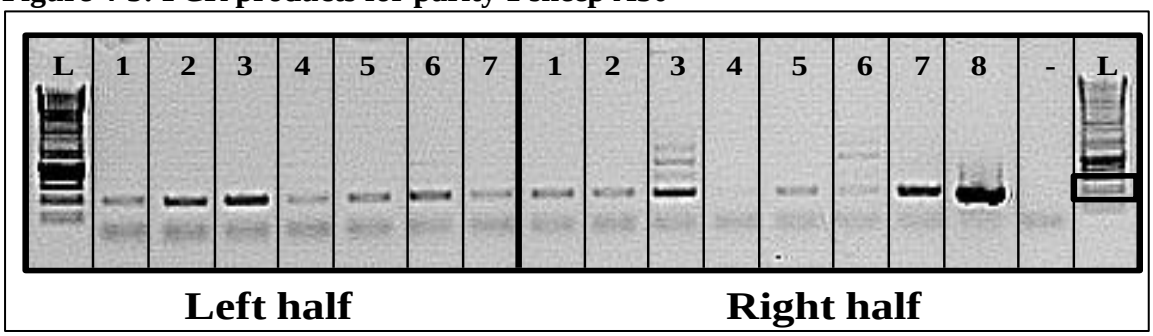


Figure 4-3: PCR products for parity 1 sheep A50



[17] In Figure 4-1 - Figure 4-3, 'L' is Hyperladder 1kb (Bioline, UK) and the 400bp marker is highlighted in the ladder on the right-hand side of the figure. Numbers 1-8 are PCR products for milk DNA for weeks 1-8, with the mammary gland half specified in the Figure. The '+' and '-' are the DNA extraction positive and negative controls respectively.

Table 4-2: Results of PCR amplification from all milk samples for Parity 1 and 2 sheep [18]

Sheep	Left half (weeks)								Right half (weeks)							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
A20	NS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	NS
A35	✓	✓	✓	✗ ✗	✓	✓	✓	NS	✓	✓	✓	✓	✗✓	NS	NS	NS
A50	✓	✓	✓	✓	✓	✓	✓	NS	✓	✓	✓	✗ ✗	✓	✓	✓	✓
A15	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗ ✗	✗ ✓	✗ ✗
A21	✓	✓	✓	✓	✓	✓	NS	NS	✓	✓	✓	✓	✓	✓	NS	NS
A26	✓	✓	✓	✓	✓	✓	NS	NS	✗ ✓	✓	✓	✓	✓	✓	✓	NS
A27	✓	✓	✓	NS	✓	✓	NS	NS	✓	✓	✓	✗✗	✓	✗✗	NS	NS
A28	✓	✓	✓	✓	✓	NS	NS	NS	✓	✓	NS	✓	✓	✓	NS	NS
A39	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
A4	✓	✓	✓	✓	✓	✓	NS	NS	✓	✓	✗ ✓	✓	✓	✓	NS	NS
A41	✓	✓	✓	✓	✓	✓	NS	NS	✓	✓	✓	✗✓	✗✓	✗✓	NS	NS
A44	✓	✓	✗✓	✓	✓	✓	NS	NS	✓	✓	✓	✓	✓	✓	✗✓	NS
A45	✓	✓	✓	✗✗	✓	✓	✓	NS	✓	✓	✗✗	✓	✓	✓	NS	NS
A46	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	NS
A47	✓	✓	✓	✓	✓	✓	✗✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
A5	✓	✓	✓	✓	✓	✓	✓	NS	✓	✓	✓	✓	✗✓	✓	✓	NS
A9	✓	✓	✗✗	✓	✓	✓	✗✗	NS	✓	✓	✓	✓	✓	✗✗	✓	NS

[18] In Table 4-2 and Table 4-3, successful amplification: ✓ and unsuccessful amplification ✗; successful amplification on a second PCR attempt: ✗ ✓; no successful amplification in two attempts (not analysed by DGGE): ✗ ✗; no sample for analysis: NS.

Table 4-3: Results of PCR amplification from milk samples for Parity 3, 4 and 10 sheep

Sheep	Left half (weeks)								Right half (weeks)							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
A12	✓	✓	✓	NS	✓	✓	✓	NS	✓	✓	✓	✓	✓	✓	✓	NS
A2	✓	✓	✓	NS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
A22	✓	✓	✓	✓	✓	✓	✓	NS	✓	✓	✓	✓	✓	✓	✓	NS
A25	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	NS	✓	✓	NS
A29	✓	✓	✓	✓	✓	✓	✓	NS	✓	✓	✓	✓	✓	✓	✓	NS
A3	✓	✓	✓	✓	✓	NS	NS	NS	✓	✓	✓	✓	✓	✓	NS	NS
A16	✓	✓	✓	✓	✓	✗✓	✓	NS	✓	✓	✓	✓	✓	✓	✓	NS
A24	✓	✓	✗✓	✓	✓	✗✗	NS	NS	✓	✓	✓	✓	NS	✗✓	NS	NS
A40	✓	✓	✓	✓	✓	✓	NS	NS	✓	✓	✓	✓	✓	✓	✓	NS
A43	NS	✓	✓	✓	✓	✓	✓	✓	NS	✓	✓	✓	✓	✓	✓	NS
A49	✓	✓	✓	✓	✓	✓	✓	NS	✓	NS	NS	✓	✓	✓	✓	NS
A6	✓	✓	✓	✓	✓	✓	NS	NS	✓	✓	✓	✓	✓	✓	NS	NS
A23	✓	✓	✓	✓	✓	✓	NS	NS	✓	✓	✓	✓	✓	✓	NS	NS

4.3.2 Denaturing Gradient Gel Electrophoresis (DGGE) analysis of PCR amplicons from milk samples

DGGE images were produced from the PCR amplicons (Section 4.3.1) for each of the 379 milk samples. There were 30 DGGE images, 1 per sheep. Exemplars of each parity are presented in this Chapter; remaining DGGE images are in Appendix 3.

4.3.2.1 DGGE results from parity 1 sheep

Three parity 1 sheep had their bacterial community visualised by DGGE. Figure 4-4 - Figure 4-6 illustrate the DGGE results for these sheep.

Figure 4-4: DGGE image of PCR amplicons for parity one sheep A20 [19]

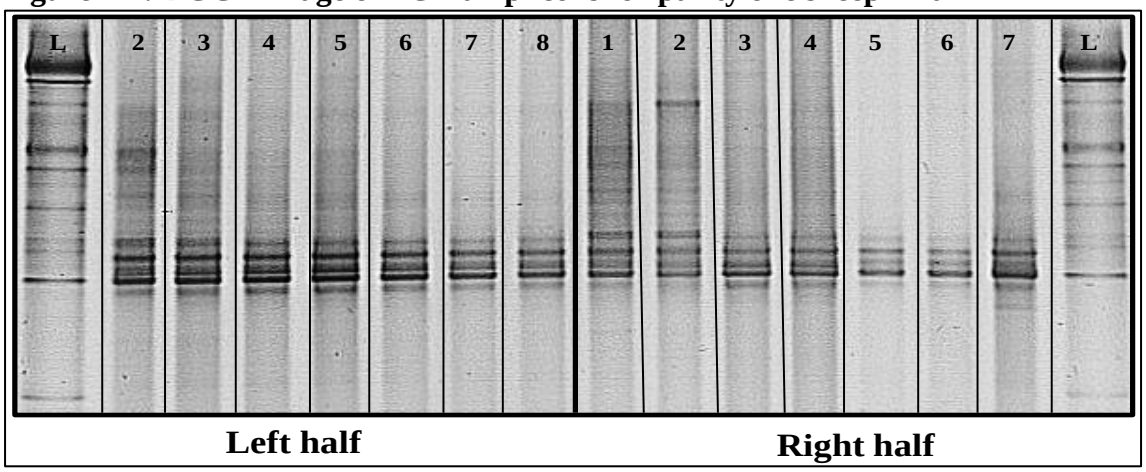


Figure 4-5: DGGE image of PCR amplicons for parity one sheep A35

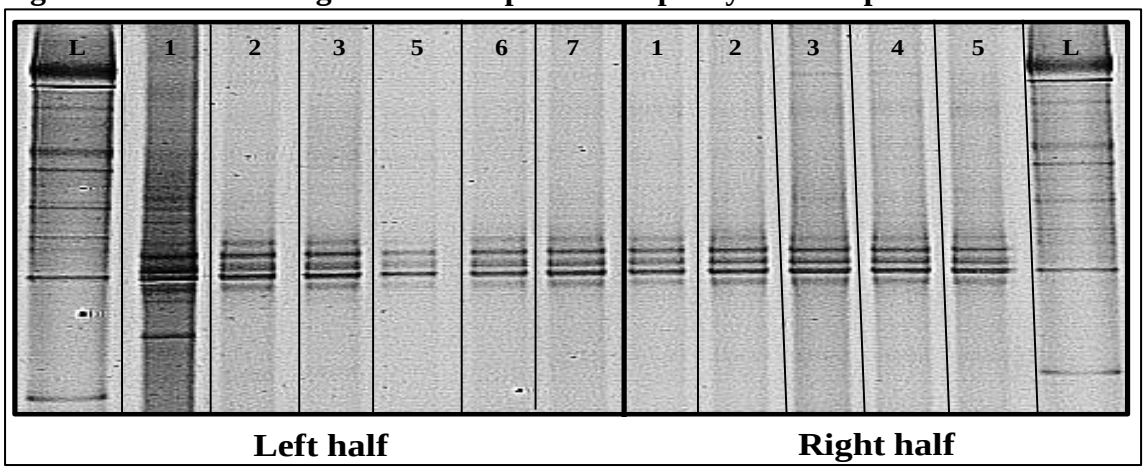


Figure 4-4 and Figure 4-5 show both similarities and differences in DGGE banding pattern within and between the two mammary gland halves of sheep A20 and A35. There is a central region of approximately 5-6 DGGE bands which appear in the majority of the samples, which

[19] For all DGGE images, 'L' corresponds to the custom DGGE reference ladder; numbers 1-8 are milk samples from weeks 1-8, with the mammary gland half they originate from specified in each Figure.

could suggest a stable community with commensal organisms persisting over lactation. However, there were also differences in banding pattern for samples from different halves at the same time point (Figure 4-5, lane 1 for both halves) suggesting community changes.

Figure 4-6: DGGE image of PCR amplicons for parity one sheep A50

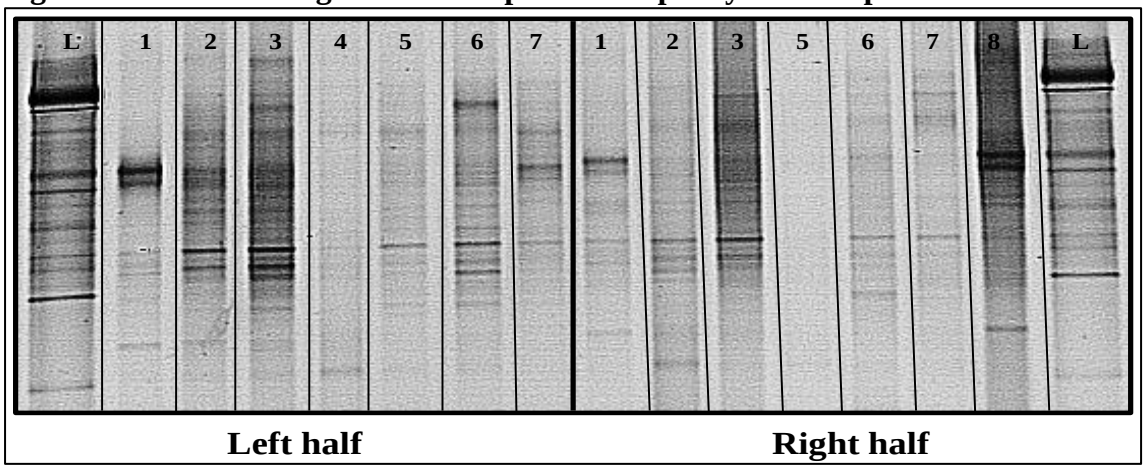


Figure 4-6 shows a more variable DGGE banding pattern in comparison to Figure 4-4 and Figure 4-5. There appears to be more DGGE bands per sample, and the bands are more widely distributed. Again, correlations can be seen between samples (Figure 4-6, lane 1 for both halves) as well as differences (Figure 4-6, lane 6 for both halves).

4.3.2.2 DGGE results from parity 2 sheep

A DGGE image was produced for each of the 14 parity 2 sheep included in the study. Figure 4-7 and Figure 4-8 are two exemplars. The other 12 parity 2 DGGEs are in Appendix 3.

Figure 4-7: DGGE image of PCR amplicons for sheep A41

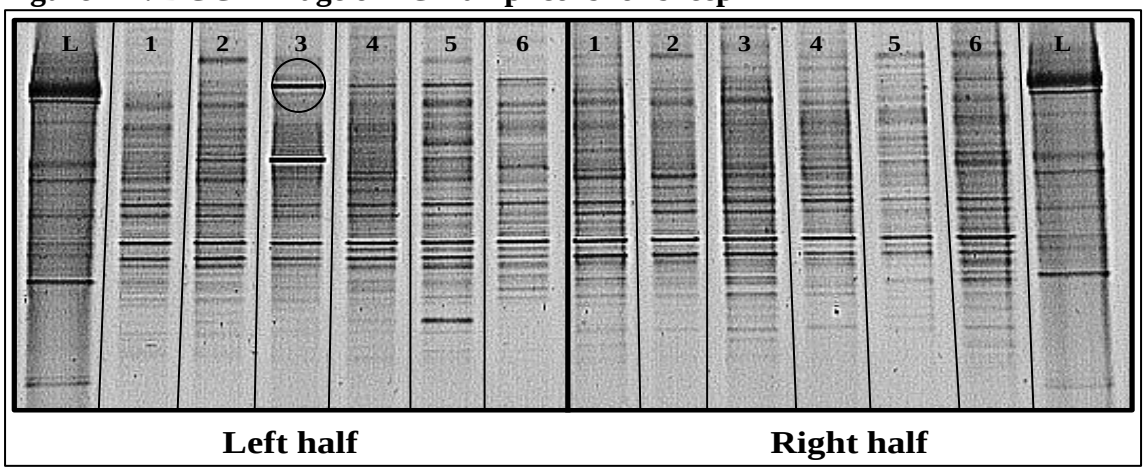
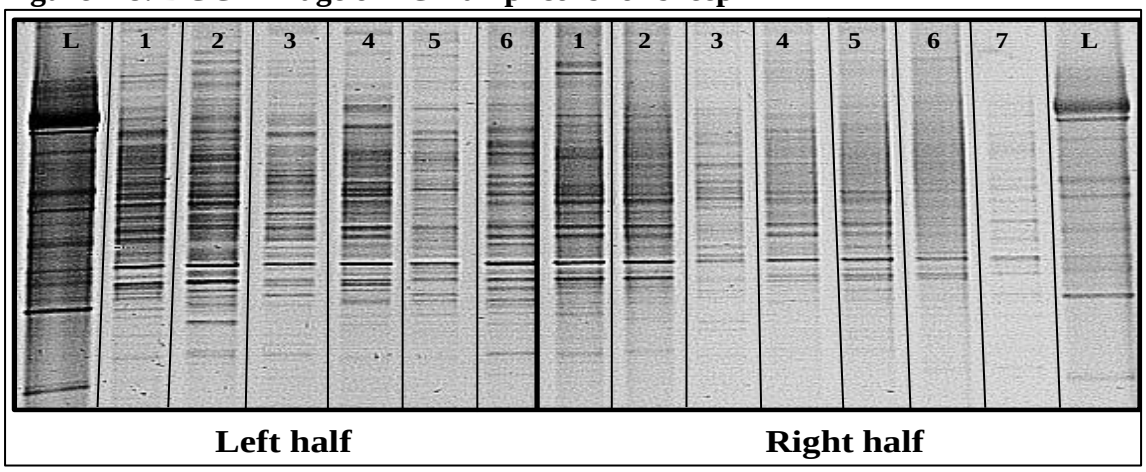


Figure 4-8: DGGE image of PCR amplicons for sheep A44



Both Figure 4-7 and Figure 4-8 show the most complex DGGE banding patterns seen in the study. Both contain multiple bands in each sample widely distributed across the DGGE gel. In Figure 4-7, there are noticeably defined bands in some samples (Figure 4-7, lane 3 of left half circled). The intensity of these bands could be related to their abundance. Figure 4-7 shows some similarities in banding pattern (Figure 4-7, lane 2 of both halves) as well as differences (Figure 4-7, lane 6 of both halves). In Figure 4-8, the intensity of DGGE banding profiles is greater in the left half, with more visible bands in comparison to the right half. This could suggest that the abundance of certain bacterial species is greater in the left half.

4.3.2.3 DGGE results from parity 3 sheep

DGGE images were produced for 6 parity 3 sheep. Figure 4-9 and Figure 4-10 are two exemplars. The DGGE images for the other 4 parity 3 sheep are in Appendix 3.

Figure 4-9: DGGE image of PCR amplicons for sheep A2

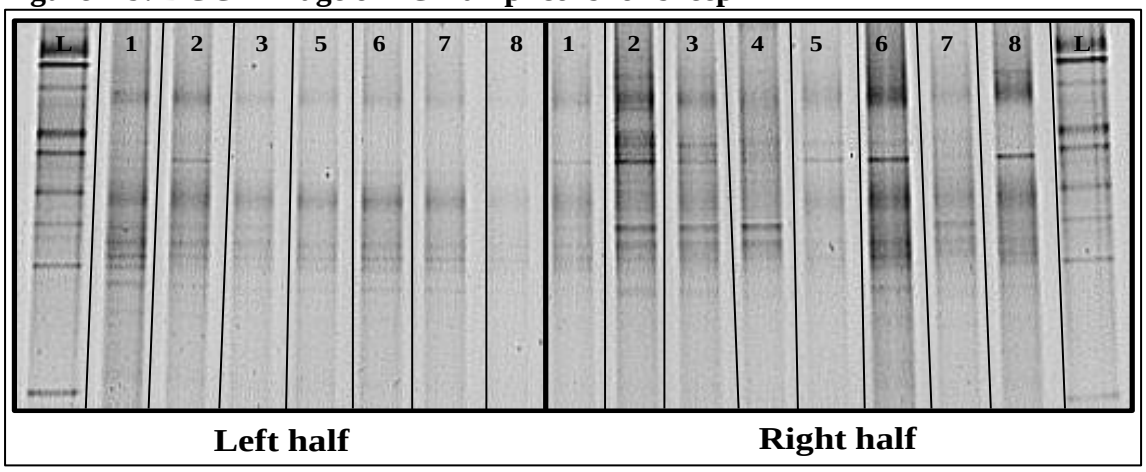


Figure 4-10: DGGE image of PCR amplicons for sheep A25

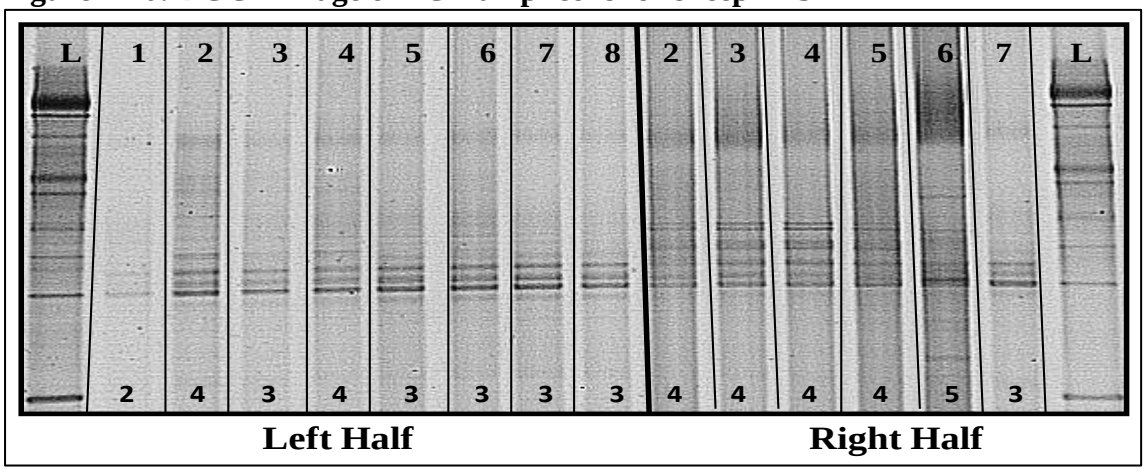


Figure 4-9 shows differences in DGGE banding pattern between left and right halves, with the right half having more bands per sample than the left half. Figure 4-10 also shows a difference in DGGE banding pattern between halves, with more bands present in weeks 2-6 of the right in comparison to the left. This correlates with the SCC data for sheep A25, which shows a higher SCC count for weeks 2-6 of the right half in comparison to the left.

4.3.2.4 DGGE results from parity 4 sheep

DGGE images were produced for 5 parity 4 sheep. Figure 4-11 and Figure 4-12 are two exemplars. The DGGE images for the other 3 parity 4 sheep are in Appendix 3.

Figure 4-11: DGGE image of PCR amplicons for sheep A40

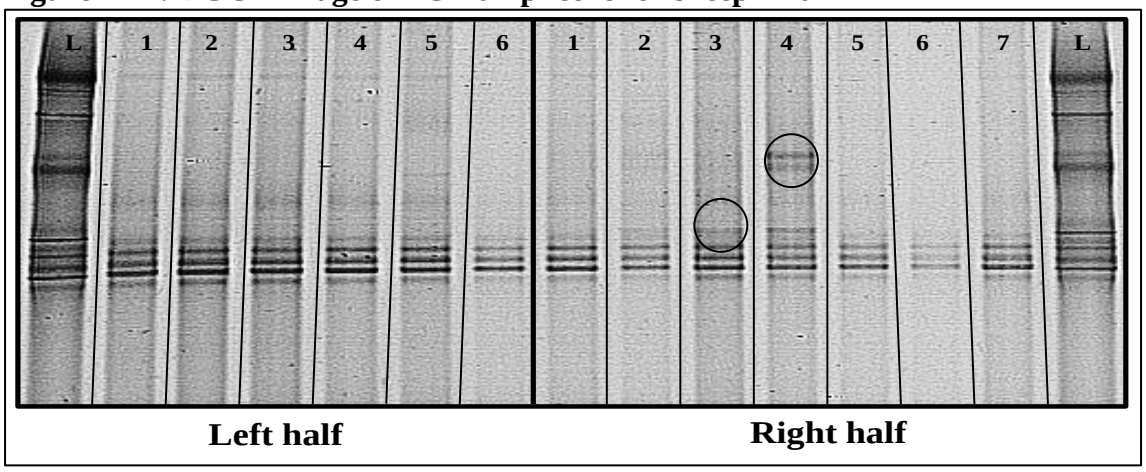


Figure 4-11 shows a consistent banding pattern of approximately 5 bands that appear in the majority of samples. Samples from weeks 3 and 4 for the right half also appear to have an additional band in this central region, with week 4 having a further 2 bands towards the top section of the Figure (circled). The sample from week 4 for the right half had a high level of

bacterial growth detected in culture-independent analysis, but had the lowest SCC score in the right half and the second lowest SCC score out of all sheep A40 milk samples.

Figure 4-12: DGGE image of PCR amplicons for sheep A49

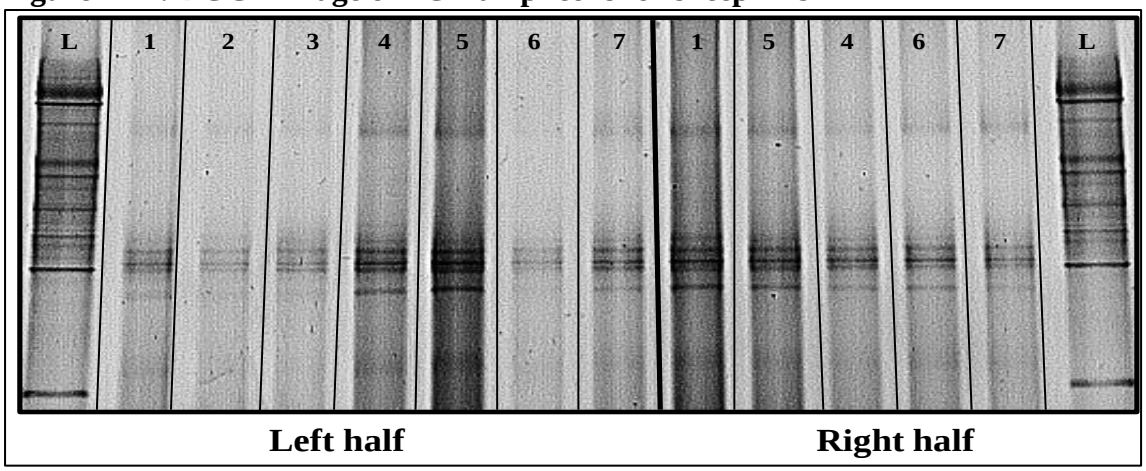


Figure 4-12 shows a consistent banding pattern across samples for sheep A49. Some similarities in banding pattern are visible e.g. Figure 4-12, lanes 4, 5 and 7 for the left half and 1, 4, 5, 6 and 7 for the right half.

4.3.2.5 DGGE results from parity 10 sheep

There were 2 parity 10 sheep. The DGGE images produced for them are shown in Figure 4-13 and Figure 4-14.

Figure 4-13: DGGE image of PCR amplicons for sheep A6

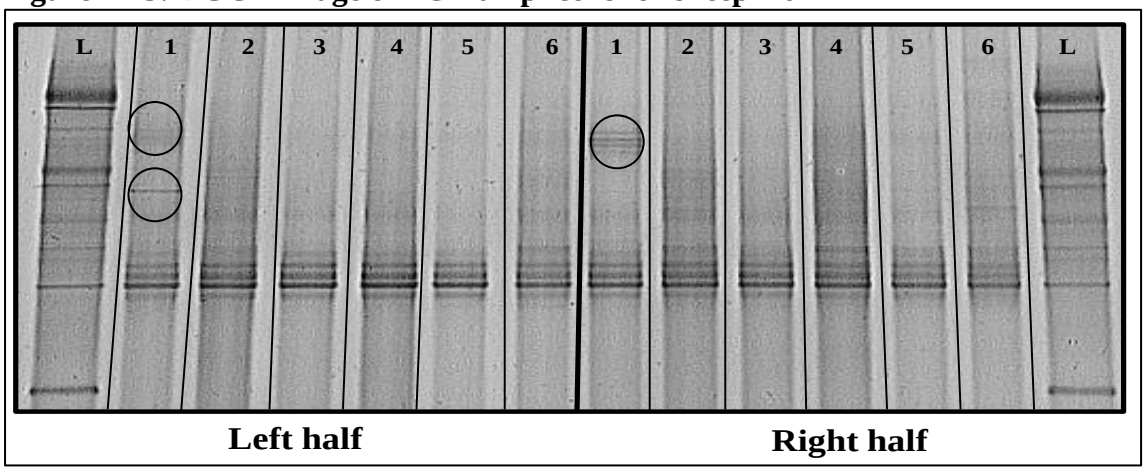


Figure 4-13 shows a similar banding pattern across samples for sheep A6 (Figure 4-13, lanes 2-6 for the left half and 2-5 for the right half). Week 1 for both halves had at least one DGGE band not seen in any other sheep A6 samples (circled).

Figure 4-14: DGGE image of PCR amplicons for sheep A23

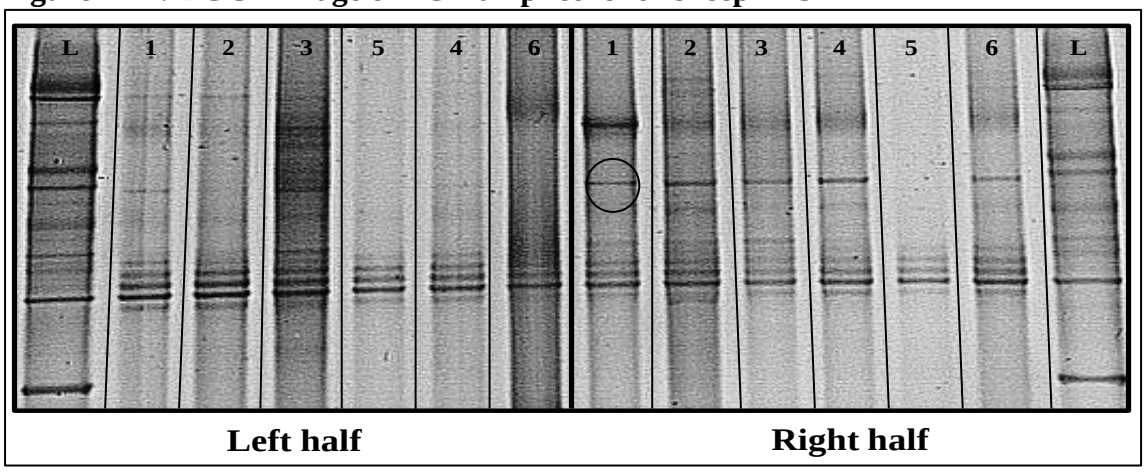


Figure 4-14 shows differences in banding pattern between the left and right halves for sheep A23 samples. The right half samples have at least 1 additional DGGE band not seen in many of the left half samples (example circled). Week 5 samples for both halves have the least visible number of bands and these 2 samples had the lowest SCC's for sheep A23 milk samples.

4.3.3 Results of Gel Compar II analysis

4.3.3.1 DGGE band count data results

The distribution of DGGE bands by parity are presented in Table 4-4.

Table 4-4: DGGE band count data per sheep ordered by parity

Sheep	Parity	No. of milk samples	Total No. of DGGE bands	DGGE band count range per sample	Mean DGGE band count (SE)
A20	1	14	89	6-8	6.4 (0.20)
A35		11	59	4-10	5.4 (0.47)
A50		13	65	2-9	5.0 (0.62)
A15	2	14	46	3-10	4.2 (0.75)
A21		12	49	1-8	4.1 (0.58)
A26		13	60	3-14	4.6 (0.78)
A27		9	62	5-10	6.9 (0.66)
A28		10	69	4-11	6.9 (0.81)
A39		16	92	2-11	5.8 (0.70)
A4		10	100	2-11	6.7 (1.02)
A41		11	115	2-23	11.2 (0.59)
A44		13	105	2-16	8.1 (1.20)
A45		11	51	2-7	4.6 (0.56)
A46		15	83	3-8	5.5 (0.49)
A47		15	62	1-10	6.2 (0.80)
A5		14	85	2-14	6.1 (0.91)
A9		11	58	2-10	5.3 (0.93)
A12	3	13	52	1-13	4.0 (1.39)
A2		15	63	1-7	4.2 (0.48)
A22		14	74	4-8	5.3 (0.30)
A25		14	49	3-4	3.5 (0.20)
A29		14	74	4-7	5.3 (0.22)
A3		11	51	4-6	4.6 (0.20)
A16	4	14	59	1-8	4.2 (0.54)
A24		10	43	1-6	4.3 (0.58)
A40		13	65	4-7	5.0 (0.20)
A43		13	101	7-10	7.8 (0.26)
A49		12	46	2-5	3.8 (0.27)
A6	10	12	64	4-7	5.3 (0.33)
A23		12	77	3-9	6.4 (0.48)
Total	**-**	**379**	**2068**	**-**	**-**
Mean	**-**	**12.6**	**68.9**	**-**	**5.5**

Table 4-5: DGGE band count data grouped according to sheep parity

Parity	No. of sheep	DGGE band range	Mean no. of DGGE bands
1	3	2-10	5.6 (0.27)
2	14	1-14	6.0 (0.23)
3	6	1-13	4.4 (0.20)
4	5	1-10	5.1 (0.25)
10	2	3-9	5.9 (0.31)
Total	**30**	**-**	**-**
Mean	**-**	**-**	**5.4**

Table 4-5 shows parity 2 and 10 sheep had the highest mean number of DGGE bands, with parity 3 and 4 sheep having the lowest. The DGGE band range was similar for most sheep regardless of parity, although parity 10 sheep had a lower maximum number of DGGE bands per sample on average which is counteracted by a higher minimum number.

Table 4-6: DGGE band data grouped according to week

Week	No. of milk samples	No. of DGGE bands	Mean no. of DGGE bands (SE)
1	56	390	7.0 (0.35)
2	59	376	6.4 (0.35)
3	56	308	5.5 (0.33)
4	53	290	5.5 (0.32)
5	56	249	4.5 (0.25)
6	52	237	4.6 (0.29)
7	35	160	4.6 (0.37)
8	12	58	4.8 (0.66)
Total	**379**	**2068**	**-**

When the data were grouped by week of lactation (Table 4-6), the mean number of DGGE bands decreased from weeks 1-4 and then marginally increased from weeks 5-8. Hence, there is a potential fluctuation in the bacterial community, with variation from week to week. When the data were grouped by mammary gland half (Table 4-7), the mean number of DGGE bands were similar.

Table 4-7: Mean number of DGGE bands grouped by mammary gland half.

Mammary gland half	Mean no. of DGGE bands (SE)
Left	5.6 (0.19)
Right	5.9 (0.17)

In the Gel Compar II analysis, every DGGE band detected in a milk sample was identified and assigned a position on the DGGE gel. Bands were then grouped according to their position. This produced 35 distinct bands and each assigned a numerical value to represent a

position on the gel. These data were coded as binary present/absent and as an intensity value between 0-255 arbitrary units (A.U.).

Table 4-8 summarises, for each band position on the DGGE gel, the number of milk samples found at that position and of these, which sheep and mammary gland half they were present in. In Table 4-8, the most common band positions were (in decreasing order); 6.98, 6.16, 6.54, 6.86, 7.13 and 6.37, which were present in 135 - 261 of the 379 milk samples. The least common band positions were (in increasing order); 0.49, 7.38, 0.98, 6.72, 2.19, 2.68, 3.84 and 7.72; all of which were present in 9 or fewer milk samples.

Table 4-9 shows a numerical measure of intensity for each of the 35 DGGE band positions. The intensity value ranges from 1 in band position 4.00 to 244 in band position 6.98. The 10 band positions with the most intense DGGE bands were (in increasing intensity); 5.50, 6.54, 4.00, 3.01, 5.68, 6.16, 7.13, 6.25, 6.37 and 6.98. The 10 band positions with the least intense DGGE bands were (in decreasing intensity); 3.01, 2.54, 6.25, 6.16, 6.86, 6.54, 5.24, 4.69, 7.13 and 4.00. The band positions 6.25, 6.37 and 6.98 had the biggest difference in band intensities between milk samples.

Table 4-8: Number of DGGE bands per band class by mammary gland half and presence in sheep [20]

Band position	No. of samples	Left half	% Left half	Right half	% Right half	No. of sheep	% of sheep
6.98	**261**	**143**	**55**	**118**	**45**	**27**	**90**
6.16	**260**	**127**	**49**	**133**	**51**	**29**	**97**
6.54	**217**	**115**	**53**	**102**	**47**	**24**	**80**
6.86	**179**	**97**	**54**	**82**	**46**	**20**	**67**
7.13	**177**	**97**	**55**	**80**	**45**	**21**	**70**
6.37	**135**	**68**	**50**	**67**	**50**	**16**	**53**
4.00	89	40	45	49	55	23	77
6.25	75	39	52	36	48	13	43
7.27	72	37	51	35	49	13	43
3.01	67	31	46	36	54	18	60
5.50	63	43	68	20	32	17	57
4.99	52	22	42	30	58	17	57
2.47	47	20	43	27	57	19	63
2.61	37	17	46	20	54	10	33
2.54	34	18	53	16	47	7	23
4.69	34	18	53	16	47	9	30
7.34	28	15	54	13	46	9	30
4.26	25	12	48	13	52	13	43
5.68	25	17	68	8	32	9	30
7.58	25	15	60	10	40	6	20
8.91	23	15	65	8	35	3	10
5.24	22	18	82	4	18	6	20
4.64	18	8	44	10	56	9	30
2.01	16	10	63	6	38	8	27
2.72	16	8	50	8	50	8	27
7.86	16	13	81	3	19	4	13
4.83	10	5	50	5	50	4	13
2.19	7	4	57	3	43	5	17
2.68	7	4	57	3	43	2	7
3.84	7	1	14	6	86	4	13
7.72	7	2	29	5	71	4	13
0.98	6	4	67	2	33	5	17
6.71	6	1	17	5	83	3	10
7.38	3	1	33	2	67	1	3
0.49	2	0	0	2	1	1	3
% of bands per half	-	**52.5%**		**47.5%**	-	-	-

[20] Table ordered according to most common DGGE band position; 6 most common band positions highlighted in bold.

Table 4-9: Number and intensity value of DGGE bands per band position [21]

Band class	No. of bands present	Max value	Min value	Mean value	S.E of mean
6.98	261	244	11	86.9	3.09
6.16	260	182	5	49.2	1.83
6.54	217	150	4	49.7	1.65
6.86	179	129	4	51.5	1.70
7.13	177	182	3	57.9	1.98
6.37	135	233	11	66.8	2.14
4.00	89	152	1	35.8	1.04
6.25	75	225	6	64.1	1.86
7.27	72	122	6	44.2	1.03
3.01	67	153	6	41.0	1.07
5.50	63	149	8	46.0	1.20
4.99	52	126	9	33.4	0.75
2.47	47	107	13	36.3	0.71
2.61	37	112	15	36.5	0.62
2.54	34	129	6	35.8	0.72
4.69	34	93	4	32.6	0.58
7.34	28	88	11	33.7	0.51
4.26	25	86	9	33.2	0.49
5.68	25	165	12	39.9	0.65
7.58	25	75	13	42.0	0.58
8.91	23	80	6	32.6	0.46
5.24	22	80	4	26.7	0.37
4.64	18	137	15	40.7	0.53
2.01	16	85	11	39.2	0.46
2.72	16	137	14	47.3	0.62
7.86	16	52	12	30.0	0.34
4.83	10	119	12	50.6	0.50
2.19	7	68	10	28.8	0.24
2.68	7	59	9	32.4	0.25
3.84	7	72	16	37.1	0.30
7.72	7	46	27	35.6	0.25
0.98	6	113	17	44.5	0.35
6.71	6	62	15	43.8	0.30
7.38	3	34	21	29.7	0.14
0.49	2	28	25	26.5	0.10
Total bands	**2068**	-	-	-	-

[21] The intensity value is an arbitrary unit (A.U.). All DGGE images were saved as 8 bit. As a result, the intensity value varies from 0 (white) to 255 (black).

4.3.3.2 SCC data

In Table 4-10, the SCC in the longitudinal study varies from 10^4-10^7 (log SCC 4-7), with the majority of sheep having SCCs in the 10^4-10^6 range. The SCC data for each sheep by parity presented in Table 4-11 indicates that the oldest sheep (parity 10) in the study had the highest mean SCC. Parity 1 and 3 sheep have the same mean SCC, followed by parity 2 and then parity 4. A higher SCC in parity 10 sheep could suggest they are most susceptible to disease, possibly due to conformational changes in the udder with age.

Table 4-10: SCC data by parity

Sheep	Parity	No. samples	Min logSCC	Max logSCC	Mean logSCC (SE)
A20	1	14	4.88	6.06	5.61 (0.098)
A35		11	4.72	5.69	5.20 (0.099)
A50		13	4.68	6.03	5.21 (0.109)
A15	2	14	4.45	6.19	5.20 (0.130)
A21		12	4.73	6.27	5.26 (0.140)
A26		13	4.30	5.65	5.00 (0.106)
A27		9	3.48	5.75	5.17 (0.223)
A28		10	4.75	5.53	5.15 (0.085)
A39		16	4.34	5.57	4.96 (0.092)
A4		10	5.06	5.95	5.39 (0.084)
A41		11	4.15	5.70	4.91 (0.133)
A44		13	4.60	5.80	5.19 (0.099)
A45		11	4.82	6.35	5.30 (0.139)
A46		15	4.38	6.06	5.13 (0.118)
A47		15	4.70	5.62	5.12 (0.071)
A5		14	5.26	5.99	5.57 (0.051)
A9		11	4.60	5.42	5.04 (0.097)
A12	3	13	4.72	5.88	5.25 (0.088)
A2		15	5.22	6.49	5.77 (0.105)
A22		14	4.88	5.53	5.18 (0.052)
A25		14	4.30	7.55	5.55 (0.285)
A29		14	4.62	5.87	5.14 (0.104)
A3		11	4.38	6.01	5.11 (0.176)
A16	4	14	4.15	5.87	5.05 (0.127)
A24		10	4.42	5.50	5.00(0.122)
A40		13	4.15	5.64	4.99 (0.125)
A43		13	4.45	5.73	5.09 (0.096)
A49		12	4.78	5.65	5.05 (0.070)
A6	10	12	4.42	5.80	5.03 (0.126)
A23		12	5.93	7.08	6.55 (0.092)
Mean	**-**	**-**	**4.61**	**5.94**	**5.24**

Table 4-11: Summary of SCC data (minimum, maximum and mean SCC) by parity

Parity	No. of sheep	Minimum SCC	Maximum SCC	Mean SCC (SE)
1	3	4.68	6.06	5.35 (0.066)
2	14	3.48	6.35	5.17 (0.032)
3	6	4.30	7.55	5.35 (0.067)
4	5	4.15	5.87	5.04 (0.048)
10	2	4.42	7.08	5.79 (0.175)

Table 4-12: Minimum, maximum and mean SCC by week

Week	No. of milk samples	Minimum SCC	Maximum SCC	Mean SCC (SE)
1	56	4.75	6.73	5.35 (0.056)
2	59	4.15	7.08	5.16 (0.077)
3	56	3.48	6.59	5.04 (0.076)
4	53	4.26	6.49	5.15 (0.068)
5	56	4.34	6.55	5.25 (0.062)
6	52	4.38	7.55	5.36 (0.062)
7	35	4.42	6.10	5.38 (0.071)
8	12	4.97	6.49	5.55 (0.141)

Table 4-12 shows the SCC decreased each week until week 5 where it started to increase. This correlates with the DGGE data which shows the number of DGGE bands decreased until week 6 where the mean DGGE band count started to rise again (Table 4-6).

Table 4-13: SCC data by mammary gland half

Mammary gland half	Minimum SCC	Maximum SCC	Mean SCC (SE)
Left	4.15	6.96	5.20 (0.034)
Right	3.48	7.55	5.28 (0.042)

As with the DGGE band count data in Table 4-7, the means are similar between mammary gland halves, although the range in SCC was greater for the right half.

4.3.3.3 GelCompar II dendrogram and Principal Component Analysis (PCA) results

The DGGE band data indicated a persistent bacterial community with at least 1 DGGE band detected in every sample. Some potential variation in both DGGE band count and SCC were seen by sheep parity and week (Table 4-5, Table 4-6 and Table 4-11 - Table 4-12) and very little variation between mammary gland half (Table 4-7 and Table 4-13). This suggested a consistent bacterial community was present, but with minor as opposed to major fluctuations in bacterial population occurring in the study period. This was further substantiated by additional analysis in GelCompar II in which within DGGE gel (i.e. within sheep) comparisons were made using a ranked Pearson Product-moment correlation coefficient and Unweighted Pair Group Method with Arithmetic mean (UPGMA). For each sheep in the study, a dendrogram and PCA plot were produced to visualise correlations between samples

within sheep, examples of these for 2 sheep are given below with the associated DGGE images included for reference; dendrogram and PCA data for all sheep are provided in Appendix 4.

Figure 4-15: DGGE results for sheep A35 [22]

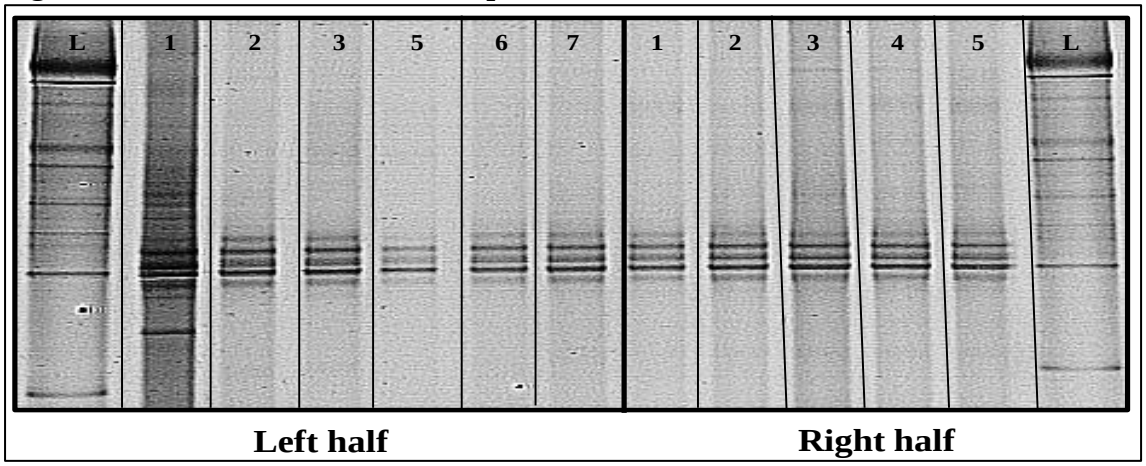
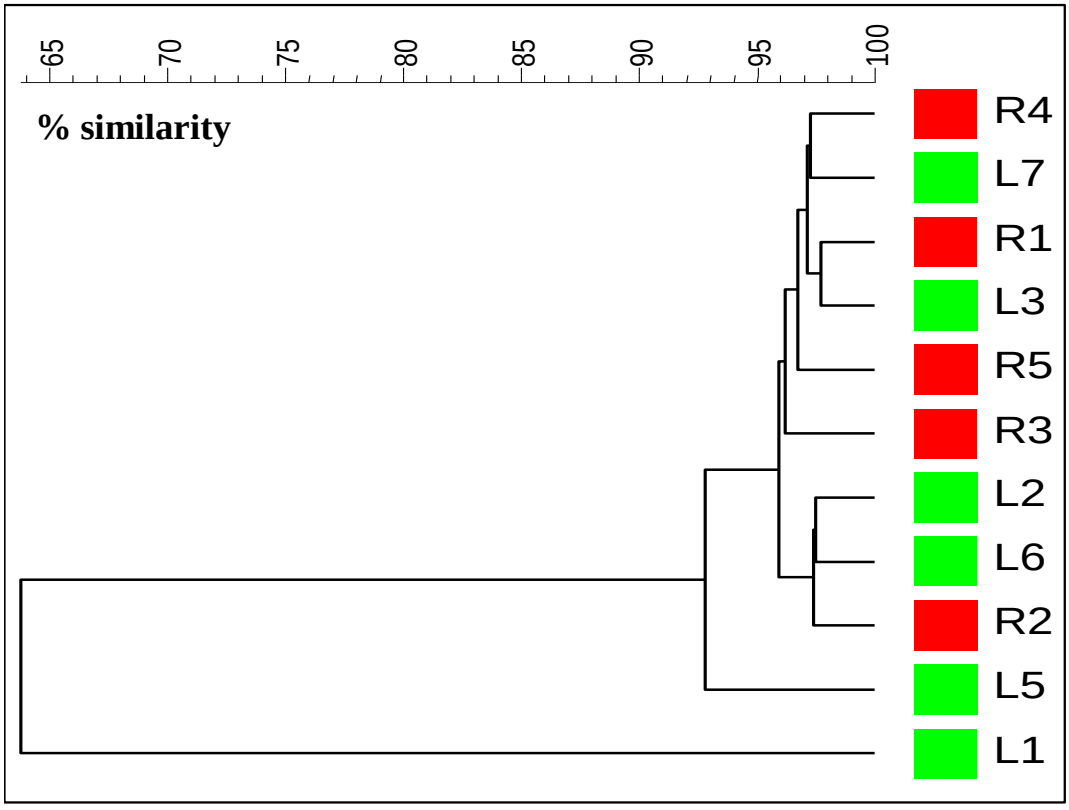


Figure 4-16: Dendrogram of sheep A35 milk samples



[22] In Figure 4-15 - Figure 4-20, each milk sample is labelled according to mammary gland half i.e. 'L' for left and 'R' for right; the subsequent number is the week i.e. 1-8; samples from the left half are green and from the right half, red. Dendrogram axis is the percentage similarity; the lowest percentage is the greatest similarity between the 2 most dissimilar samples. The axis on the PCA plots are the entry coordinates which are the coordinates of the entries in the first 2 components of the PCA.

Figure 4-17: PCA plot of sheep A35 milk samples

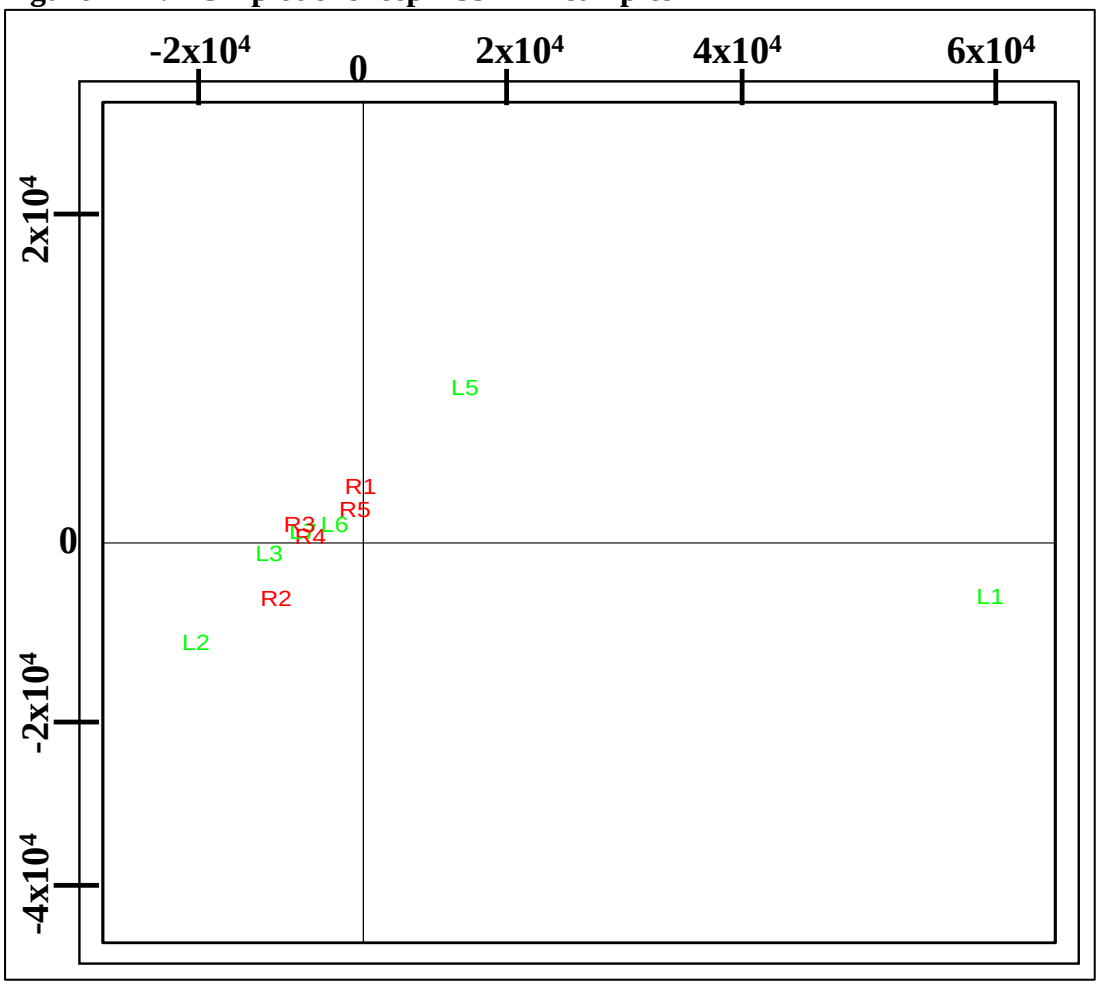


Figure 4-18: DGGE results for sheep A25

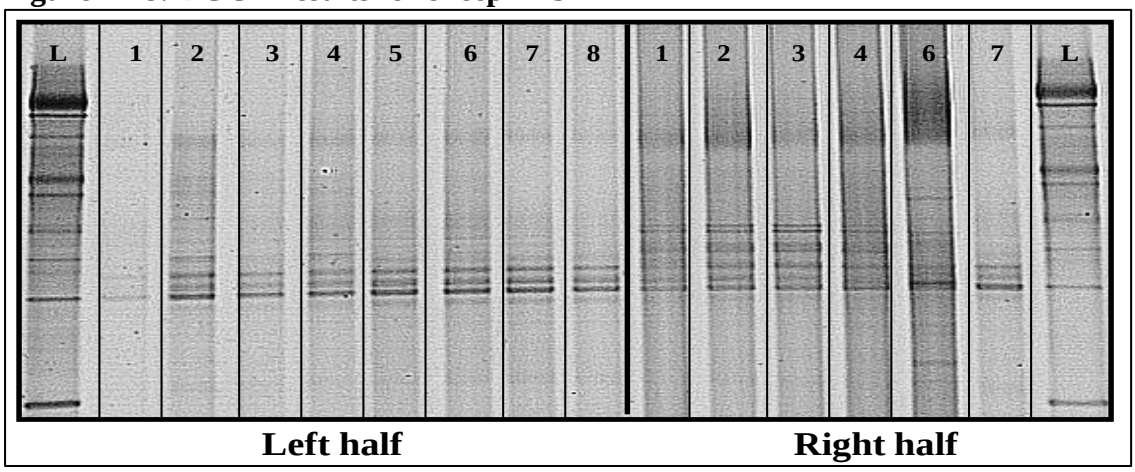


Figure 4-19: Dendrogram of sheep A25 milk samples

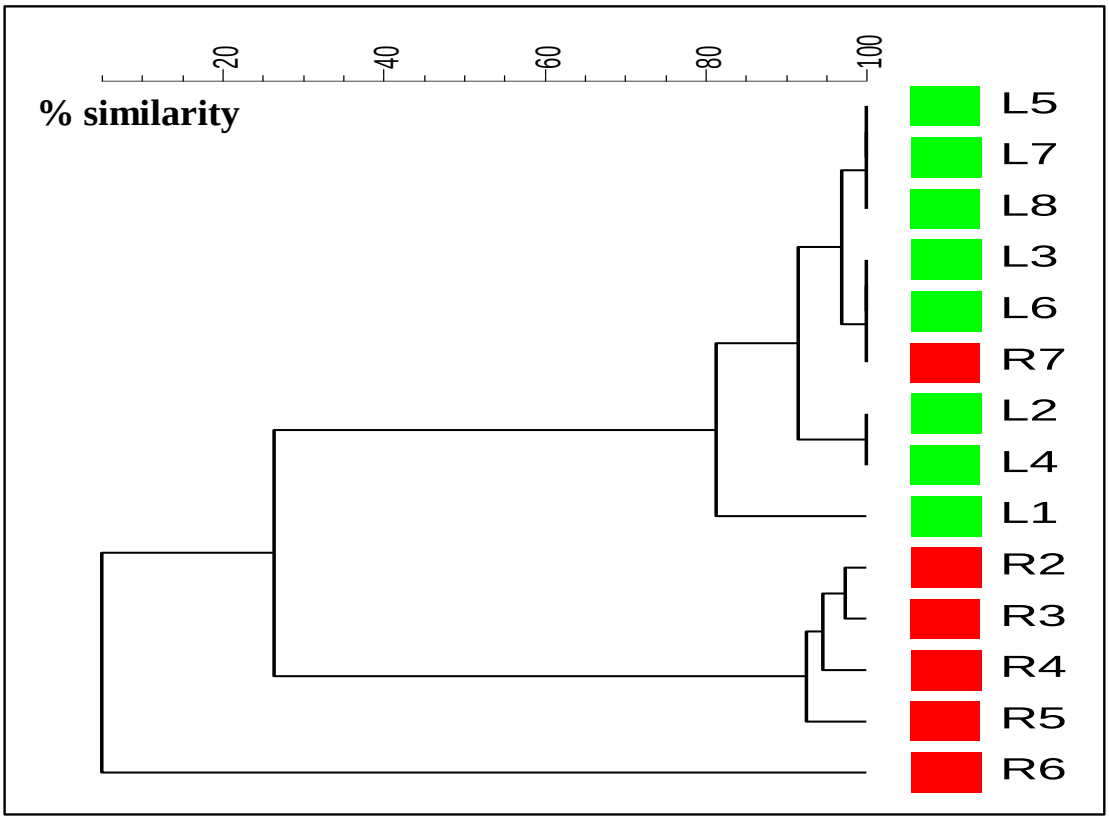


Figure 4-20: PCA plot of sheep A25 milk samples

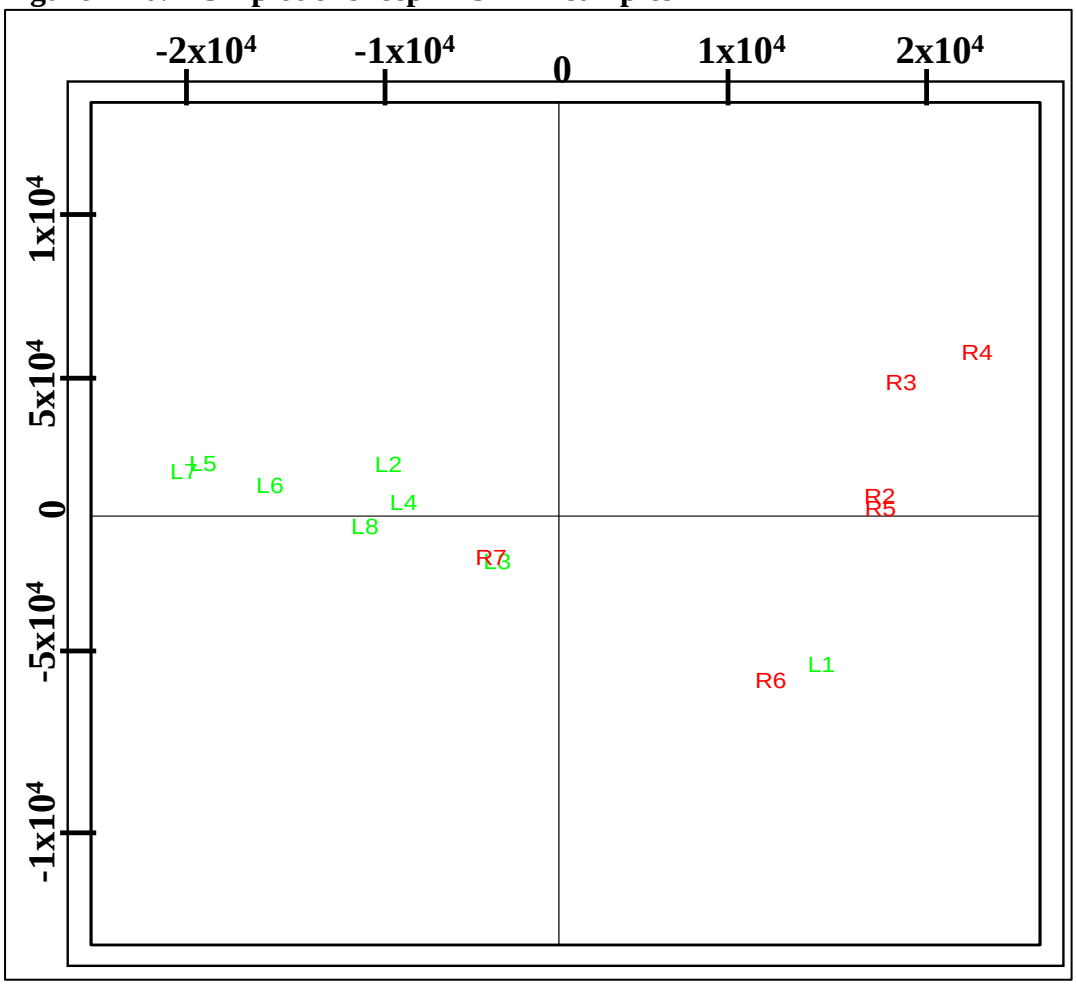


Figure 4-16 - Figure 4-20 show the dendrogram and PCA plots for sheep A35 and A25. For sheep A35, there was no discernible pattern with all but two samples (L1 and L5) outliers of a tight cluster of the rest of the samples. This is also seen in the PCA plot for sheep A35 (Figure 4-17) where the majority of the samples are grouped in the same region, with no pattern according to mammary gland half.

The dendrogram for sheep A25 (Figure 4-19) shows more of a clustering pattern according to mammary gland half, with left and right half samples clustering separately (apart from R7 which clusters with left half samples). In Figure 4-20, the division of the majority of the samples according to mammary gland half for sheep A25 can also be seen.

The results from these two sheep are indicative of the results seen across sheep in the study i.e. despite some difference in DGGE banding pattern within and between sheep, samples within a sheep tended to either cluster together or by mammary gland half or a combination of both. Some similarity between halves could be expected as all of the sheep are from the same farm and the left and right halves are in close proximity to one another on each individual sheep. Differences would also be expected as well the halves are separate entities.

A similarity matrix of the bacterial communities in all milk samples was analysed using non-metric multidimensional scaling (non-metric MDS) and analysis of similarity (ANOSIM) in PRIMER, version 6, Primer-E (Clarke and Gorley, 2006). Data were clustered by sheep (Figure 4-21), parity (Figure 4-22) and week of lactation (Figure 4-23 and Figure 4-24) to investigate the effect of each factor on community similarity between individuals and over time.

There was no significant clustering of total number of DGGE bands by sheep, suggesting that there are differences in bacterial community diversity between sheep (Figure 4-21). Early (1 and 2) and late (3 and 4) parities formed two clusters, suggesting differences in bacterial community diversity according to sheep age (Figure 4-22). There were no clear patterns discernible when investigating all samples individually over time (Figure 4-23), however analysis of the mean band count by SCC per week revealed differences in bacterial community diversity according to week. Earlier weeks clustered more closely to each other than later weeks, indicating a change in bacterial community diversity over the sampling time.

Figure 4-21: Non-metric MDS plot of all study data clustered by sheep.

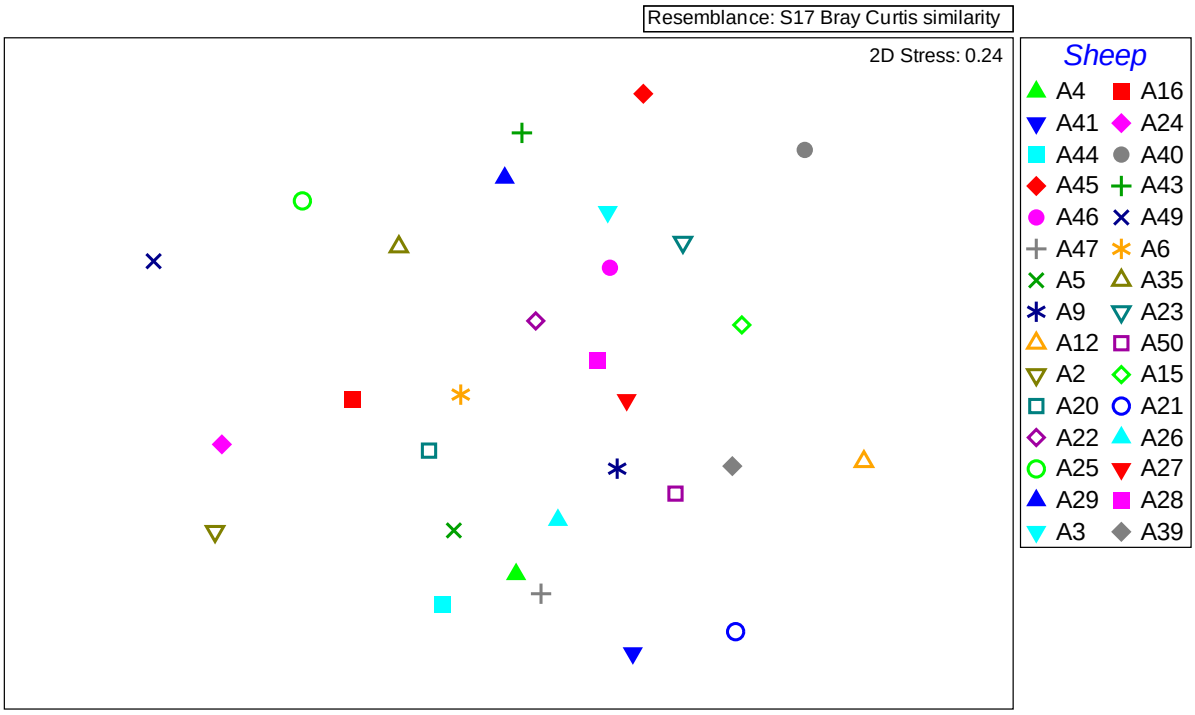


Figure 4-22: Non-metric MDS plot of all study data clustered by sheep parity (age).

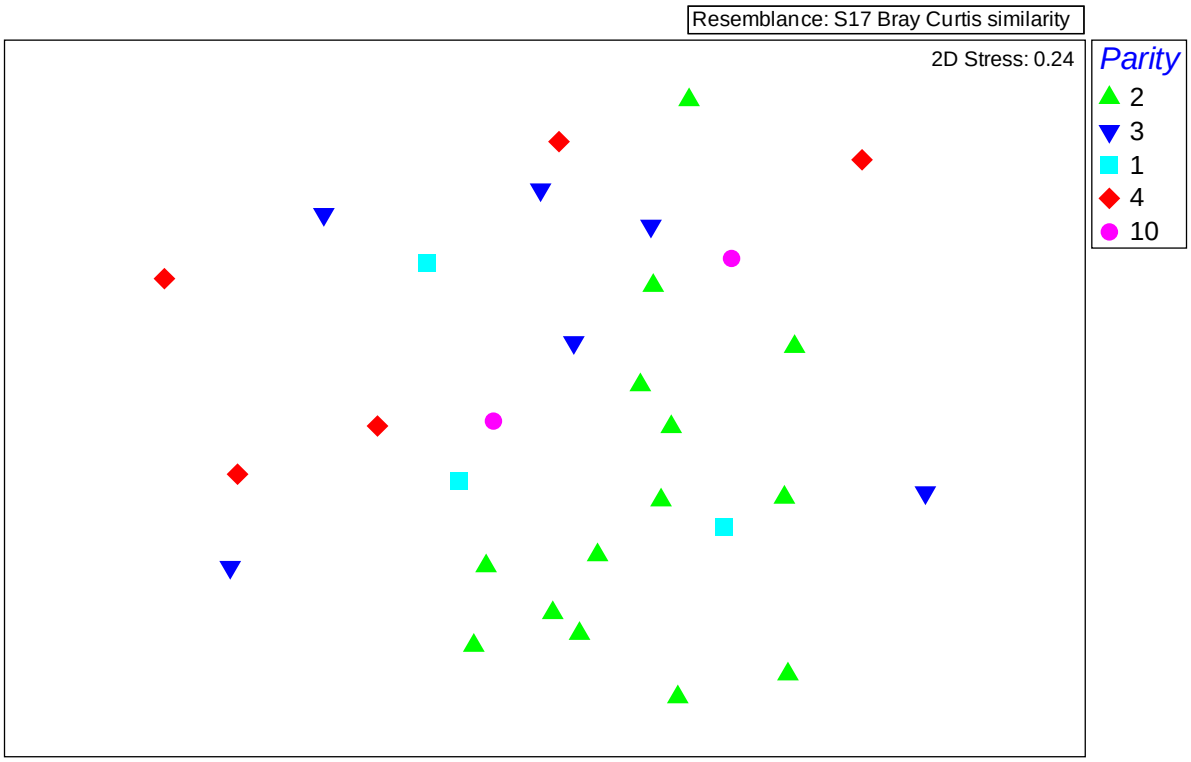


Figure 4-23: Non-metric MDS plot of all study data clustered by week of lactation (1-8).

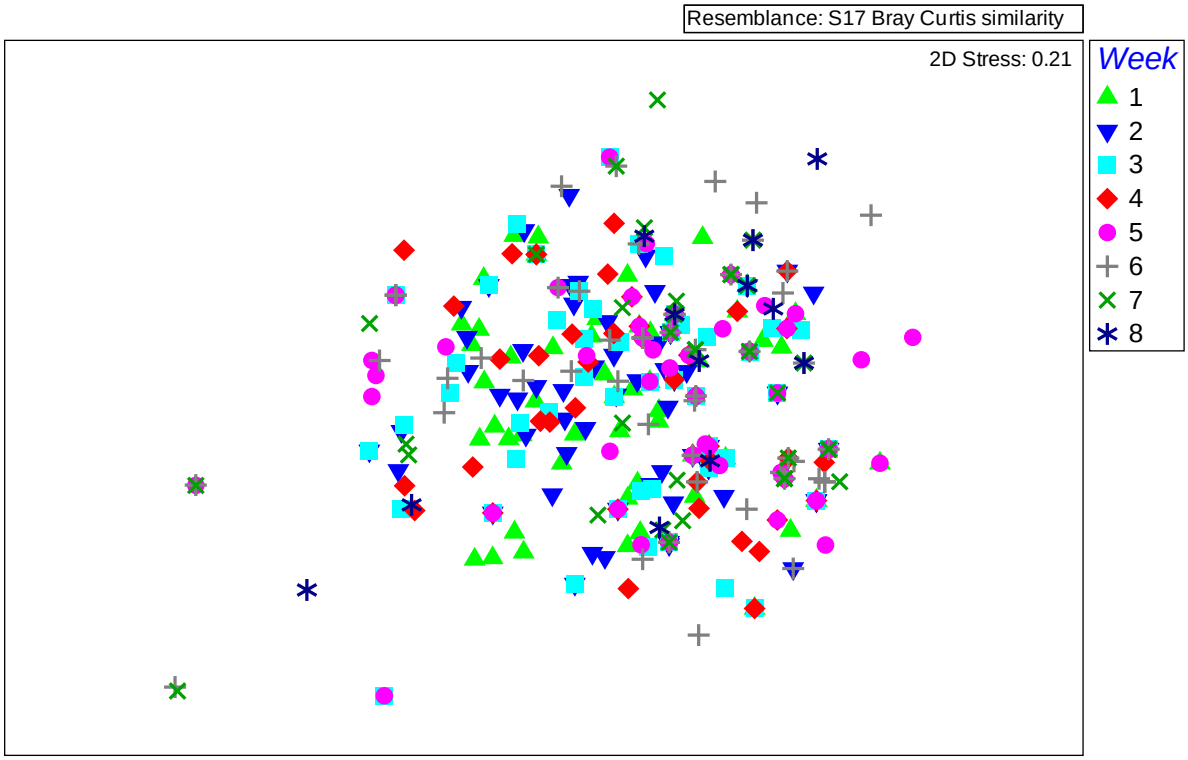
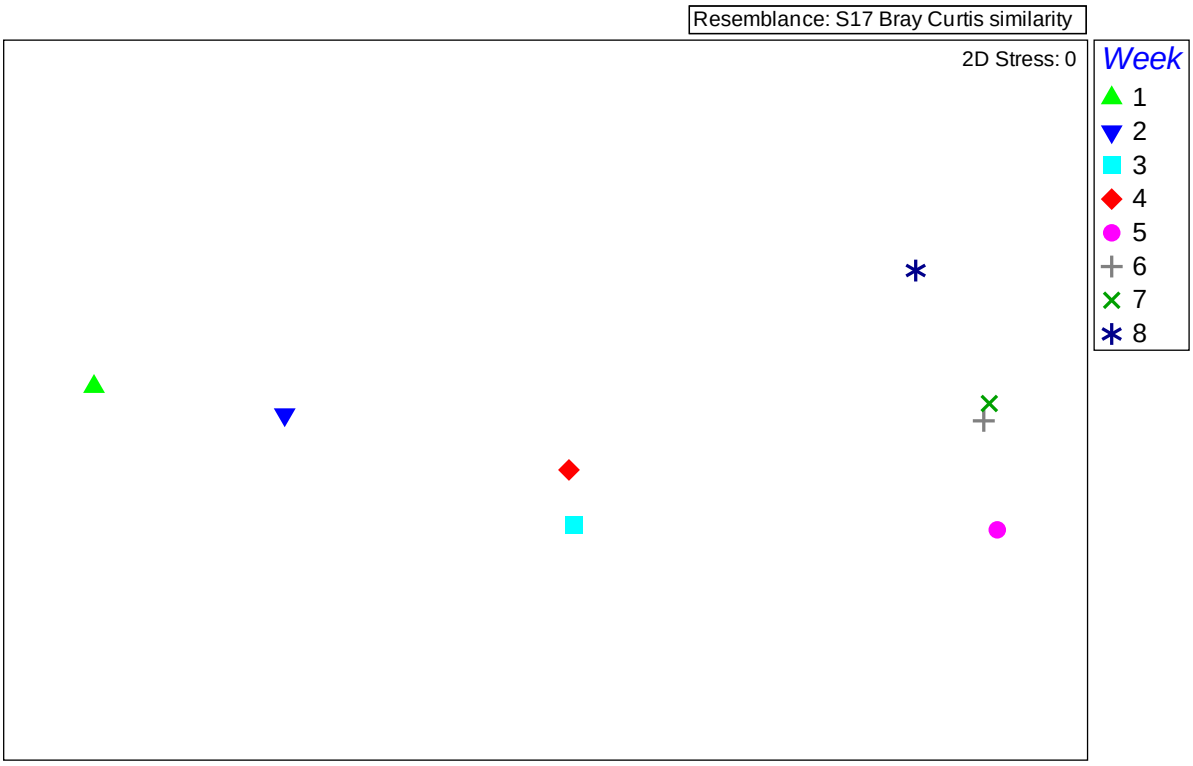


Figure 4-24: Non-metric MDS plot of all study data clustered by week of lactation (1-8) using the mean DGGE band count and LogSCC per week.



The ANOSIM analysis of the study data supports the indications from the non-metric MDS plots, with sheep, parity and week of sampling shown to have a significant effect on the bacterial community diversity (Table 4-14). The non-metric MDS plot grouping samples by mammary gland half indicated no clustering by half (not shown) and was supported by the ANOSIM (Table 4-14).

Table 4-14: ANOSIM results of all study data analysed by sheep, parity, mammary gland half and week of lactation [23]

Factor	P-value	R-value
Sheep	**0.001**	0.661
Parity	**0.002**	0.081
Mammary gland half	0.093	0.005
Week	**0.002**	0.028

The non-metric MDS and ANOSIM results indicated that investigating community composition in individual sheep over time in further detail was required to identify specific bacterial species associated with a change in disease state (SCC).

[23] P-values showing significance of effects of factors by which the data was grouped (sheep, parity, mammary gland half and week). R-values close to zero indicate most similarity. Values in bold highlight significant differences ($p = <0.05$).

4.3.4 Modelling of GelCompar II data

A total of 15 of the 35 DGGE band positions were significantly associated with change in SCC. To identify the bacterial species associated with each band, a selection of milk samples from different sheep, mammary gland halves and weeks were re-analysed by DGGE. Significant bands were excised and either sequenced directly or cloned before sequencing. The DGGE band identifications are shown in Table 4-14.

Table 4-15: Identifications for 15 DGGE bands that had a significant effect on SCC [24]

							Identified from:	
Band	**SCC effect**	**No. of milk samples present in**	**ID**	**Phyla**	**BLAST identity (%)**	**No. IDs**	**No. milk sample**	**No. sheep**
6.16	↑	260	*Corynebacterium efficiens*	A	94	5	4	2
2.54	↑	34	*Psychrobacter maritimus*	P	99	3	1	1
2.01	↑	16	*Streptococcus uberis*	F	99	1	1	1
8.91	↑	16	*Burkholderia cepacia*	P	98	2	1	1
2.19	↑	7	*Fusobacterium necrophorum*	Fu	98	9	2	2
3.84	↑	7	*Trueperella pyogenes*	A	99	3	1	1
7.72	↑	7	*Pseudomonas chlororaphis*	P	99	13	3	2
0.49	↑	2	*Psychrobacter faecalis*	P	98	2	2	2
6.54	↓	217	*Achromobacter xylosoxidans*	P	99	3	1	1
6.37	↓	135	*Rhodococcus qingshengii or Nocardia globerula*	A	98	15	10	5
3.01	↓	67	*Atopostipes suicloacalis*	F	94	4	2	2
4.69	↓	34	*Mannheimia haemolytica*	P	92	2	1	1
2.68	↓	7	*Jeotgalicoccus psychrophilus*	F	96	5	2	2
7.38	↓	3	*Sharpea azabuensis*	F	99	2	2	1

[24] In 'Phyla' column, P: Proteobacteria; F: Firmicutes; Fu: Fusobacteria; A: Actinobacteria.

As shown in Table 4-14, 8 DGGE bands were associated with a higher SCC and 6 associated with a lower SCC (Table 4-14). DGGE band position 7.86 was the fifteenth DGGE band and was significantly associated with a lower SCC. However, no visible PCR product was produced post-excision from three repeat DGGE gels and so band 7.86 could not be identified.

4.4 Discussion

4.4.1 Using the PCR-DGGE approach

PCR-DGGE was chosen as an effective method to gain a snapshot of the bacterial community because it has been used previously to improve the diagnosis of the etiological agents of mastitis and enable the identification of different bacterial species involved, including uncultivable isolates (Chen and Hwang, 2008; Kuang *et al.*, 2009). For example, Kuang *et al.,* (2009) excised 26 DGGE bands from 4 mastitic cow milk samples for identification and nearly 50% were related to uncultured clones. Despite identifying only coliforms from on-farm via culturing, Kuang *et al.,* (2009) found that *Lactococcus lactis* and *Staphylococcus aureus* were widely distributed using PCR-DGGE. Delgado *et al.,* (2008) identified DGGE-specific bands from human breast milk samples not detected by culturing. Braem *et al.,* (2012) excised 78 bands and successfully identified 81% of these from 48 teat apex samples.

Density changes in DGGE bands can also give an indication of relative abundance. For example, Delgado *et al.,* (2008) cultured *Staphylococcus aureus* isolates from milk samples most frequently and *Staphylococcus aureus* were also the most intense DGGE band.

Hence, as shown both in published literature and this study, PCR-DGGE offers a reliable and reproducible method to visualise community profiles and identify prominent community members (Muyzer and Smalla, 1998).

4.4.2 Bacterial identification with PCR-DGGE

A range of bacteria has been identified in milk using PCR-DGGE. Braem *et al.,* (2012) identified 4 bacterial phyla (*Actinobacteria, Bacteroidetes, Firmicutes* and *Proteobacteria*) encompassing 17 bacterial genera from 48 samples from dairy cow teat apices, the findings of which have been corroborated by other studies (Kuang *et al.*, 2009; Rasolofo *et al.*, 2010). Human skin microbiome studies are also in agreement with these findings. For example, Costello *et al* (2009) found 92% of species were from the same four phyla when investigating samples from different sites of the human body. The results from this study support these

previous findings with four bacterial phyla (Table 4-14) found to have a significant effect on SCC as discussed further in Section 4.4.4.

Results from literature and this study suggest that there is a commensal microbial community in milk. For example, Braem *et al.,* (2012) reported that highest number of bacterial genera recovered from teat apices of dairy cow quarters were from those considered un-infected. Other authors have reported a wide diversity of commensal organisms in milk including *Aerococcus, Acinebacter, Bacteroidetes, Proteobacteria, Corynebacterium* and coagulase-negative staphylococci as well as known pathogens such as *Staphylococcus aureus* and *Escherichia coli* (Braem *et al.*, 2012; Giannino *et al.*, 2009; Kuang *et al.*, 2009; Rasolofo *et al.*, 2010). In the current study, a minimum of 1 DGGE band was detected in each of the 379 milk samples processed using the PCR-DGGE approach, suggesting a community was present in all milk samples, regardless of SCC.

4.4.3 Interpretation of DGGE band count and SCC data

DGGE patterns from milk are often unique and of varying complexity with reports of between 2 and 19 prominent DGGE bands per sample, with additional faint bands (Delgado *et al.*, 2008; Delgado *et al.*, 2013; Giannino *et al.*, 2009; Kuang *et al.*, 2009). Specific DGGE banding patterns have been reported between farms, hosts and cow quarters (Braem *et al.*, 2012; Delgado *et al.*, 2008; Giannino *et al.*, 2009; Kuang *et al.*, 2009).

In this study, 1-23 DGGE bands were identified per milk sample. However, this could be an overestimate of diversity, because bacteria can harbour more than one copy of the 16S rRNA encoding gene, with heterogeneous sequences, giving rise to more than one band (Muyzer and Schäfer, 2001; Nubel *et al.*, 1996). Alternatively, an underestimation of diversity is possible, because only bacterial populations that make up more than 1% of the total community can be detected by DGGE and dissimilar sequences can co-migrate to the same position in a DGGE gradient (Muyzer and Schäfer, 2001; Rossello-Mora *et al.*, 1999). However, potential extra bands should not interfere with conclusions based on comparison of patterns from different samples (Muyzer and Schäfer, 2001). Despite this, MiSeq analysis of amplified 16S rRNA genes from 64 milk samples identified between 3-17 OTUs per milk sample, indicating some similarity to the DGGE band counts.

Similarities and differences were seen between halves, within halves, over time and between sheep. For some sheep, complex DGGE profiles were present that comprised several bands corresponding to a wide variety within the bacterial population e.g. Figure 4-7, whilst for

others, more consistent and simple profiles were seen e.g. Figure 4-12. Some samples did not produce sufficient PCR product for clear visualisation of DGGE banding profiles e.g. Figure 4-6, lane 5 of right half. This highlights the difficulty in analysing non-disease samples, where bacterial loads may be lower than the limit of detection for the methods currently available.

Despite this, there were also DGGE band intensity changes e.g. intensity difference between and left and right halves in Figure 4-8, which could suggest changes in the abundance of certain community members within a half and/or over time. In some cases there were intensity changes but the DGGE banding pattern remained constant e.g. Figure 4-12. Increases in intensity empirically corresponded to an increase in SCC e.g. Figure 4-12, lanes 4 and 5 of the left half have the two highest SCC scores for this half and produce the most intense bands. This could suggest the community remains stable, but there are fluctuations in abundance of some community members which could be linked to a shift towards a disease state.

However, this was not always the case as some samples had an increase in SCC which was not linked to a more complex DGGE profile. For example, in Figure 4-11, two additional DGGE bands appeared in lane 4 of the right half, yet this sample has the lowest SCC score of all the samples in this half and the second lowest SCC score for this sheep overall, despite lane 4 being the only sample in this sheep with presumably high bacterial growth (Appendix 1). As PCR-DGGE has detection limits and not every band is a single bacterial species as previously discussed (Muyzer and Schäfer, 2001), it is possible that samples for this sheep have a more complex composition than appears on the DGGE image. Alternatively, the additional bands in lane 4 could represent new bacterial species entering the community which are not having a significant effect on SCC.

Both DGGE and SCC data showed some variation by parity and week of lactation. For parity, the highest mean SCC and second highest mean number of DGGE bands occurred in parity 10 sheep i.e. the oldest sheep in the study. Parity has been associated with an increased risk of severe clinical mastitis in dairy cattle (Biffa *et al.,* 2005; Green *et al.,* 2007; Green *et al.,* 2002; Peeler *et al.,* 2002). Increasing SCC without disease in dairy cattle and sheep (Green *et al.,* 2005; Huntley *et al.,* 2012; Reneau, 1986) and conformational changes in the udder could predispose to infection (Huntley *et al.*, 2012). It has also been suggested that infections in previous lactations may damage/alter the mammary environment or reduce innate defence

mechanisms (Green *et al.*, 2005) or that there could be a greater response in a previously infected half (Dohoo and Meek, 1982; Macmillan *et al.*, 1983) or an increased number/level of detection of persistent infections with age (Fthenakis, 1996).

An increased risk of infection with increasing parity provides some evidence for a persistent bacterial community developing, with changes over time acting as triggers for infection. Interestingly, all of the parities below 10 had similar mean SCC and DGGE band counts. It could be hypothesized from this that the younger sheep have a more rapidly developing mammary microbiome as they are exposed to more bacterial species for the first time in comparison to older sheep. In turn, it could be hypothesized that parity 3 and 4 sheep have more stable communities after exposure in earlier years, before deterioration of the mammary environment and/or defence mechanisms in older age lead to an increase in SCC. It must be noted that low SCC is not always an indicator of good health, as dairy cattle with very low SCCs (<40,000 cells/ml) are more susceptible to severe clinical mastitis. However, $\geq$92% of the milk samples in this study had SCCs exceeding this value.

SCC decreased for up to the first five weeks of lactation and then marginally increased for the remaining weeks. Fluctuations in SCC over time have been reported for cow bulk tank milk (Peeler *et al.*, 2002; Schukken *et al.*, 2003) and individual cows; Bradley and Green (2001b) reported more severe clinical mastitis in early lactation for dairy cattle in 6 herds. It is interesting to note that, in the first 8 weeks of lactation, none of the sheep had a recorded episode of clinical disease, yet 19% of milk samples had an SCC $\geq$ 400,000 cells/ml which would be considered subclinically infected (van Schaik *et al.*, 2002). As the Gel Compar II analysis suggests that community diversity is relatively similar within sheep, it could be hypothesized that changes in bacterial interactions as opposed to diversity may have a role in disease development within sheep. Community diversity is also similar across some sheep, although some sheep did have more complex DGGE profiles, which suggests there may also be subject-specific microbiomes; a finding in human breast milk studies (Hunt *et al.*, 2011).

No significant difference in the number of DGGE bands was found between mammary gland halves. Some similarities between halves could be expected as transmission of the same bacterial strains between cow quarters (Phuektes *et al.*, 2001) as well as persistence of the same bacterial pathogens before and after the development of clinical symptoms in dairy sheep have been reported (Fthenakis and Jones, 1990). However, each half is a separate entity so differences in bacterial community composition would be expected. Green *et al.*, (2002)

found evidence of differing quarter susceptibility to intramammary infection in dairy cattle, which does indicate the expectation of differences in community composition between mammary quarters/halves.

4.4.4 Identification of bacterial species significantly associated with SCC change

4.4.4.1 Bacterial species associated with higher SCC

A mixed effects logistic regression model identified 15 of the 35 DGGE band positions as having a significant effect on SCC, with 8 associated with a higher SCC and 7 with a lower SCC. Of these, 14 were identified to species level by cloning and Sanger sequencing. *Corynebacterium efficiens*, *Psychrobacter maritimus*, *Streptococcus uberis*, *Burkholderia cepacia*, *Fusobacterium necrophorum*, *Trueperella pyogenes*, *Pseudomonas chlororaphis* and *Psychrobacter faecalis* were significantly associated with a higher SCC.

Corynebacterium efficiens is a Gram-positive, non-motile, non-spore forming bacteria isolated most commonly from soil and vegetable samples (Fudou *et al.*, 2002). It has been found in several studies of bacterial communities in raw cow milk (Braem *et al.*, 2012; Raats *et al.*, 2011) as well as in high-throughput sequencing studies of cow milk (Kuehn *et al.*, 2013; Oikonomou *et al.*, 2013).

Corynebacterium species have been associated with high SCC and both an increased and decreased risk of clinical mastitis in cattle, depending upon the stage of lactation when a cow is detected first infected (Bradley *et al.*, 2007; Green *et al.*, 2002; Pankey *et al.*, 1985). That said, intramammary infection with *Corynebacterium* species at drying off increased the risk of clinical mastitis in the next lactation, yet its presence during the dry period reduced the risk (Green *et al.*, 2002). It could be that cattle with mammary glands colonised by minor pathogens such as *Corynebacterium* species during one lactation have a weaker immune system or poorer defence mechanisms, meaning they are more likely to develop clinical mastitis in the next lactation.

Psychrobacter faecalis and *Psychrobacter maritimus* are Gram-negative *Proteobacteria* originally isolated from the bioaerosol of a room contaminated with pigeon faeces (Kampfer *et al.*, 2002) and from coastal sea ice and sediments of the Sea of Japan respectively (Romanenko *et al.*, 2004). The majority of *Psychrobacter* isolates of human origin are not considered clinically relevant, although there have been cases where they have acted as opportunistic pathogens (Gini, 1990; Lozano *et al.*, 1994). Most *Psychrobacter* species

identified to date have been isolated from the environment, although *Psychrobacter* species are close phylogenetic relatives of *Moraxella* and *Acinetobacter* species, which have been implicated in human infections and are well-established animal pathogens (Vela *et al.*, 2003). Little is known about their clinical significance, but *Psychrobacter faecalis* has previously been isolated from raw cow milk (Delbes *et al.*, 2007) and other strains have been isolated from diverse habitats such as sea water and lamb carcasses (Maruyama *et al.*, 2000; Prieto *et al.*, 1992). *Psychrobacter pulmonis* was isolated in pure culture from the lungs of two lambs that died suddenly (Vela *et al.*, 2003) so some *Psychrobacter* species appear to cause opportunistic disease.

Streptococcus uberis is commonly associated with clinical mastitis in cows (Bradley *et al.*, 2007), where it has been reported to cause approximately 30% of clinical mastitis cases (Hillerton and Berry, 2005). Intramammary infection with *Streptococcus uberis* can result in subclinical, clinical and chronic infections that can reoccur or persist for more than 1 lactation (Abureema *et al.*, 2014; Oliver *et al.*, 1998). *Streptococcus* spp. were associated with an increased log10SCC in a study of dairy cattle milk (Oikonomou *et al.*, 2014). *Streptococcus uberis* has been isolated from a range of sites including milk, cow teat and udder surface, tonsils, straw bedding and mechanical milking devices (Cullen, 1966; Marogna *et al.*, 2010). Faecal contamination of the dairy environment with *Streptococcus uberis* as well as repeated infections with new strains of the bacterium, mean environmental hygiene is a key component of disease prevention when *Streptococcus uberis* is present (Abureema *et al.*, 2014).

Burkholderia cepacia (formerly *Pseudomonas cepacia*) are Gram-negative non-spore forming bacilli which have a primary habitat which includes river sediment and soil and are the causal agent of soft rot in onions (Berriatua *et al.*, 2001; Govan *et al.*, 1996). This highly diverse group of bacteria have now emerged as important opportunistic pathogens capable of causing infections particularly in human immunocompromised individuals such as cystic fibrosis patients (Govan *et al.*, 1996; Mahenthiralingam *et al.*, 2008). There have been few reports of *Burkholderia cepacia* infection in veterinary medicine. Despite this, Berriatua *et al.*, (2001) reported the *Burkholderia cepacia* complex as the causal agent in an outbreak of subclinical mastitis in a flock of 620 dairy sheep. In the study, 2-3 fold increases in bulk SCC were reported. The origin of the infection was not identified, although as no animals had recently been introduced into the flock, an environmental source was proposed to be most likely.

Fusobacterium necrophorum is a Gram-negative, non-spore forming anaerobe and opportunistic pathogen that causes necrotic conditions including liver abscesses, periodontal diseases, laryngitis and infectious foot disease in a variety of species (Antiabong *et al.*, 2013; Calvo-Bado *et al.*, 2011; Lechtenberg *et al.*, 1998; Lechtenberg *et al.*, 1988; Nagaraja *et al.*, 2005; Panciera *et al.*, 1989). The role of *Fusobacterium necrophorum* in infectious disease initiation and progression has been a frequent topic of discussion, although recent findings provide evidence of its role as a secondary invader (Witcomb *et al.*, 2014). *Fusobacterium necrophorum* is also reported to be present in mixed infections, forming synergistic relationships with other bacterial species (Panciera *et al.*, 1989; Takeuchi *et al.*, 1983). It has previously been isolated in pure culture from milk samples taken from clinically affected cow quarters and has been associated with cases of summer mastitis (Jousimies-Somer *et al.*, 1996; Madsen *et al.*, 1992; McGillivery *et al.*, 1984). Hence, *Fusobacterium necrophorum* has the potential to invade a mammary gland and cause disease.

Trueperella pyogenes, recently reclassified from the genus *Arcanobacterium* (Yassin *et al.*, 2011) is an opportunistic pathogen that can cause acute suppurative mastitis in dairy cattle known as summer mastitis as well as liver abscesses and pneumonia (Jost and Billington, 2005; Zastempowska and Lassa, 2012). It has been associated with high SCC in diseased cattle and causes the most significant milk loss alongside major mastitis pathogens such as *Staphylococcus aureus* and *Escherichia coli* (Grohn *et al.*, 2004; Malinowski *et al.*, 2006).

Pseudomonas chlororaphis was also significantly associated with a higher SCC. *Pseudomonas* species have been isolated from raw cow milk and clinical mastitis cases from human breast milk using PCR-DGGE (Delgado *et al.*, 2008; Lafarge *et al.*, 2004; Raats *et al.*, 2011; Rasolofo *et al.*, 2010). *Pseudomonas* species were also dominant in subclinically infected cow milk detected using high-throughput sequencing techniques (Bhatt *et al.*, 2012; Kuehn *et al.*, 2013). Kuehn *et al.*, (2013) found *Pseudomonas* and *Psychrobacter* species in significantly higher abundance in milk samples from cow quarters with no clinical signs of disease in comparison to clinical milk samples. This could suggest these bacterial species are commensals which dominate in subclinical conditions to cause opportunistic disease.

More than one pathogen has been found in mastitic milk samples, suggesting interactions occur between different bacterial species resulting in synergistic or inhibitory influences (Kuang *et al.*, 2009). Further correlations between the presence/absence of certain bacterial species and the subsequent risk of mastitis have been previously reported. For example,

Green *et al*., (2002) found that the probability of isolating *Escherichia coli* or *Streptococcus uberis* was significantly greater when the other organism was cultured in a milk sample. El-Khodery and Osman (2008) found a mixed infection of *Escherichia coli* and *Klebsiella pneumoniae* was more prevalent than a single infection at both the animal and quarter level. Hence, the variable effect of some bacterial species seen in the literature could be a result of interaction with other bacterial community members.

4.4.4.2 Bacterial species associated with lower SCC

Six DGGE bands were associated with a significantly lower SCC. These were; *Achromobacter xylosoxidans, Nocardia globerula* or *Rhodococcus qingshengii, Atopostipes suicloacalis, Mannheimia haemolytica, Jeotgalicoccus psychrophilus* and *Sharpea azabuensis*.

Achromobacter xylosoxidans is a Gram-negative, aerobic, motile bacterium that is an opportunistic pathogen reported to cause infections in immunocompromised human individuals (De Baets *et al.*, 2014; Tena *et al.*, 2014; Trancassini *et al.*, 2014).

Nocardia globerula is not commonly associated with mastitis, although members of the same genus have been identified in raw milk and cheese as well as in bacterial communities from water collected from areas with recent oil spillages (Delbes *et al.*, 2007; Ogino *et al.*, 2001). *Nocardia* species have been used in antibiotic production (Hoshino *et al.*, 2004; Tanaka *et al.*, 1997) so may have a role in suppressing other bacterial species and so decreasing SCC. Oikonomou *et al.*, (2014) found *Nocardiodes* to be associated with a lower SCC in a study of dairy cattle milk.

Rhodococcus qingshengii is a Gram-positive, aerobic, non-motile mesophilic strain isolated from a palm tree rhizosphere soil sample (Bala *et al.*, 2013). There is little information in the literature linking *Rhodococcus* species to mastitis, apart from a case of *Rhodococcus equi* granulomatous mastitis in an immunocompetent human patient (Nath *et al.*, 2013). *Nocardia globerula* and *Rhodococcus qingshengii* could not be distinguished from one another using the BLAST search with the 16S rRNA Bacteria and Archea database.

Atopostipes suicloacalis has only previously been identified in fed-batch composting reactors (Watanabe *et al.*, 2008).

Despite being associated with a lower SCC in this study, *Mannheimia haemolytica* is a pathogen associated with clinical mastitis in sheep (Jones, 1985) and is considered a

commensal of the tonsils of lambs and adult sheep (Biberstein *et al.*, 1970). Its prevalence varies from 21% to 92% between studies and countries (Onnasch *et al.*, 2002; Scott and Jones, 1998) and it has been found in up to 100% of healthy lamb tonsils in experimental studies (Jones and Watkins, 2000). In one study, it was responsible for approximately 40% of clinical mastitis cases in ewes suckling their lambs (Mavrogianni *et al.*, 2007).

Jeotgalicoccus psychrophilus is a Gram-positive, facultatively anaerobic bacterium originally isolated from Korean fermented seafood (Yoon *et al.*, 2003). It has been isolated from raw milk and cheese and members of the same genus have been found on the teat apex of dairy cows (Braem *et al.*, 2012; Delbes *et al.*, 2007). *Jeotgalicoccus* species have been found in a wide range of environments including sediment, soil, salt lakes and poultry houses (Guo *et al.*, 2010; Martin *et al.*, 2011).

Sharpea azabuensis is a Gram-positive strict anaerobe. It was isolated from the faeces of 4 thoroughbred horses (Morita *et al.*, 2008), but has not been associated with mastitis.

Many of the bacterial species associated with higher SCC have a known association with mastitis in both cows and sheep e.g. *Streptococcus uberis*, *Trueperella pyogenes*, *Corynebacterium* species and *Pseudomonas* species. Others such as *Fusobacterium necrophorum*, *Psychrobacter* species and *Burkholderia cepacia* are opportunistic pathogens with reports of links to mastitis in the literature. In the current study, *Mannheimia haemolytica* is a known mastitis pathogen in sheep, yet was associated with a lower SCC. Malinowski *et al.*, (2006) reported the presence of the same bacterial species in cow milk samples with high and low SCC, with only the proportions of bacterial species changing with SCC. This is supported by the findings of this study, where consistent DGGE banding patterns with changes in intensity could indicate potential changes in bacterial load. If known pathogens can persist when there are no clinical signs of disease, it could be hypothesized that other factors such as timing of infection, other bacterial species present, animal immune response and environmental conditions may play a role in the disease state of the mammary glands.

Bacteria found in a diverse range of environments but unknown as mastitis pathogens were also associated with low SCC e.g. *Jeotgalicoccus psychrophilus*, *Atopostipes suicloacalis* and *Sharpea azabuensis*, in addition to opportunistic pathogens such as *Achromobacter xylosoxidans*. It could be hypothesized that these are commensal organisms that have a role in stabilising the mammary gland microbial community. *Nocardia globerula* is not known to be

associated with mastitis, yet has antimicrobial properties, so it to could be a commensal that offers a protective effect from potentially pathogenic community members. Other commensals detected in cow and sheep milk such as *Lactococcus lactis* have been suggested to have a protective effect. For example, Kuang *et al.,* (2009) found that 1 milk sample with the most intense DGGE band for *Lactococcus lactis* had no Gram positive organisms detected which might imply its presence may have had a protective role.

The identification of anaerobes such as *Fusobacterium necrophorum*, *Jeotgalicoccus psychrophilus* and *Sharpea azabuensis* highlights the validity of using a culture-independent approach to assess communities of organisms where currently only aerobic culture is used.

In conclusion, many of the bacterial species significantly associated with a change in SCC have either been linked to intramammary infection or have acted as opportunistic pathogens found in various environmental locations in previous studies.

4.4.5 Improvements for future studies

Upon reflection, identification of representative members of all 35 DGGE bands would have provided a more extensive insight into bacterial community composition and diversity because the species not associated with changes in SCC would also have been identified. Linking all species to the modelling would have allowed the relationships between community members associated and not associated with a change in SCC to be investigated. Of particular interest would have been the identity of bacterial species not associated with a change in SCC. With 19% of milk samples having a SCC that would be considered infected in dairy cattle, it could be hypothesized that other known mastitis pathogens not detected in the sequenced bands and hence not associated with a change in SCC would be present as in a natural microbial community some bacterial pathogens would persist regardless of disease status. Further elucidation of community composition will be achieved with the sequencing of all study samples as discussed in Chapter 5.

This study was the first longitudinal study to the author's knowledge that explores the microbial community in suckler sheep mammary glands. PCR-DGGE was used as a screening tool to identify whether a community could be detected in non-diseased milk samples before progression to identification of community members. Excised bands required re-PCR and cloning before DNA sequences of sufficient quality could be obtained for identification. Kuang *et al.*, (2009) reported difficulty in identifying some DGGE bands because they could not be excised from the DGGE gel or produced no PCR amplification due

to their low intensity. Twenty-three different milk samples from 13 sheep were re-screened just to obtain less than half of the DGGE band positions, and for many of these only 1 successful identification from 1 milk sample from 1 sheep was achieved. Ideally, to be confident of identifications, the same DGGE band would have been successfully excised from several milk samples, halves and time points.

The difficulty in identifying community members using PCR-DGGE highlighted the necessity of using other methods to discern more about bacterial community composition and interactions. This led to the high-throughput sequencing of study samples which is discussed in Chapter 5.

The Gel Compar analysis in Section 4.3.3.3 indicated a consistent bacterial community over time within sheep, with samples from within a half more likely to be similar to each other than those between halves. Hence, the abundance of individual bacterial species is important to consider when investigating roles and interactions within a microbial community. The intensity of individual bands on a DGGE gel can give an indication of relative abundance as previously discussed, but only Q-PCR can provide quantitative data on specific community members. Hence, further studies would benefit from the incorporation of Q-PCR data which would facilitate comparisons between the load of pathogenic to benign or commensal bacterial strains.

4.4.6 Chapter 4 conclusions

PCR-DGGE provided evidence for a persistent bacterial community in the sheep mammary gland. Similarities and differences were seen within mammary gland halves through lactation, between halves and between sheep. SCC and bacterial community diversity appeared to have weekly fluctuations throughout lactation, but overall similar patterns within mammary gland halves. There were both similarities and differences in DGGE banding pattern within sheep, but the majority of milk samples within a sheep clustered together. There were some similarities across sheep, but also some clear differences in DGGE profiles between sheep, suggesting some had personalised microbial communities.

Opportunistic and known mastitis-causing bacterial pathogens were associated with both significantly higher and lower SCCs than average, suggesting bacterial pathogens may be present when there are no clinical signs of disease. However some milk samples had SCCs that would be considered infected, yet DGGE analysis in Gel Compar II did not identify any major differences in community diversity across samples within sheep. Even parity 2 sheep

that had complex DGGE banding patterns showing potentially more bacterial species, had similar clustering profiles to less complex DGGEs of other sheep. This indicates persistence of similar bacterial species through lactation. That said, regardless of whether milk samples were clinical or not, bacteria, including pathogens, were present in all milk samples. Hence, it could be hypothesized that bacterial load and/or interactions in the community play an important role in disease.

High-throughput sequencing of milk samples would provide a more detailed census of the bacterial community composition to further investigate the associations between the diversity and type of bacteria present and increased susceptibility to intramammary infection.

Chapter 5 : Methodology for Illumina MiSeq analysis of bacterial 16S rRNA gene in sheep milk samples

5.1 Introduction

DGGE offered a unique insight into bacterial community diversity and in association with statistical modelling, provided information on community members associated with a change in animal immune response (SCC) across sheep milk samples. However, in order to gain a more detailed consensus of community composition and to provide clarity on the associations detected between bacterial species in Chapter 4, next-generation sequencing was required.

The Illumina MiSeq platform was selected to analyse sheep milk samples. In comparison to other next-generation sequencing technologies, the MiSeq requires less DNA, has a lower base call error rate, equivalent read quality and substantially reduced costs (Caporaso *et al.*, 2012; Quail *et al.*, 2012). Illumina uses a sequencing-by-synthesis approach. This involves using fluorescently labelled reversible-terminator nucleotides on clonally amplified DNA templates immobilized to an acrylamide coating on the surface of a glass flow cell (Bentley *et al.*, 2008; Quail *et al.*, 2012).

High-throughput sequencing using 454-barcoded pyrosequencing has been used to assess bacterial diversity in dairy cow milk (Bhatt *et al.*, 2012; Kuehn *et al.*, 2013; Oikonomou *et al.*, 2014; Oikonomou *et al.*, 2012; Oikonomou *et al.*, 2013). Oikonomou *et al.*, (2012) assessed the bacterial community composition in 156 milk samples from a mixture of healthy, subclinical and clinical mastitis cases. Discriminant analysis of sequencing data showed the microbiota of samples derived from healthy quarters was different from mastitic quarters, which correlates with the findings of Kuehn *et al.*, (2013) and a later study by the same authors (Oikonomou *et al.*, 2014). Such studies suggest there are changes in community diversity that can be related to disease state.

Interestingly, Oikonomou *et al.*, (2012) also detected anaerobic species, which cannot be detected using standard microbiological culture as it uses only aerobic conditions. Anaerobic bacteria detected included *Fusobacterium necrophorum* and *Trueperella pyogenes* in mastitic milk samples, with a later study finding a low prevalence of anaerobic organisms in healthy

and subclinical milk samples (Oikonomou *et al.*, 2014). These findings are supported by the results from Chapter 4 in this study in which both *Fusobacterium necrophorum* and *Trueperella pyogenes* were associated with a significantly higher SCC.

Kuehn *et al.,* (2013) compared culture-negative mastitic and non-mastitic dairy cow milk samples from the same animal. The genera *Brevundimonas, Burkholderia, Sphingomonas* and *Stenotrophomonas* were significantly associated with clinical milk samples and *Pseudomonas, Psychrobacter* and *Ralstonia* with healthy milk samples. *Burkholderia* spp. *Pseudomonas* spp. and *Psychrobacter* spp. were all associated with a significantly higher SCC in Chapter 4. It could be hypothesized that different species within these genera have differing affects on SCC.

Oikonomou *et al.,* (2014) assessed the bacterial diversity in 150 dairy cows, with SCCs ranging from healthy to clinical status. The study found milk samples had a diverse microbial community regardless of their SCC status. Similarly, *Staphylococcus* spp. and *Streptococcus* spp. including *Streptococcus uberis* were the most prevalent in all milk samples, despite their association with clinical mastitis (Section 1.2.2). Hunt *et al.,* (2011) also found *Staphylococcus* spp. and *Streptococcus* spp. dominated healthy human breast milk. These findings suggest the presence of a commensal microbial flora which includes bacterial pathogens and so it is reasonable to hypothesize that changes in the balance and/or interaction of different community members may trigger disease; the DGGE profiles in Chapter 4 also indicate the presence of several bacterial species per milk sample, including bacterial pathogens identified in Section 4.3.4, regardless of SCC.

Similarly, Bhatt *et al.,* (2012) identified 56 different bacterial species of varying abundance and mastitis pathogens including *Escherichia coli* and *Staphylococcus aureus* dominating in subclinical dairy cattle milk samples. *Proteobacteria* and *Firmicutes* were the main phyla identified in the milk samples; two-thirds of the bacterial species associated with a change in SCC in Chapter 4 are from these two phyla.

Hunt *et al.,* (2011) characterised the bacterial community in the breast milk of 16 women, collecting 3 samples over a 4 week period. The bacterial community was often, but not always, stable over time, with some communities relatively unchanging over time, whilst others had shifts in the relative abundance of bacterial genera. This links to the idea suggested in Chapter 4 of a change in DGGE profile intensity and banding pattern in some sheep milk samples potentially correlating to shifts in the abundance of certain species and changes in

community diversity. The same study found the most abundant genera to be: *Streptococcus*, *Staphylococcus*, *Serratia*, *Pseudomonas* and *Corynebacteria*. *Streptococcus*, *Pseudomonas* and *Corynebacteria* were associated with a change in SCC in Chapter 4.

Interestingly, Hunt *et al.,* (2011) also identified a 'core' microbiome of 9 OTUs in every sample from every human subject, although the relative abundance of these OTUs varied greatly between samples. Due to their dominance, these 'core' OTUs could have a protective role in the mammary gland, although 50% of the community was not conserved across the 16 women, suggesting communities associated with individuals are often highly personalised. There is no reason to suggest this is any different for sheep, with differences in DGGE profiles across sheep suggesting animal-specific microbial community compositions.

The sheep milk samples in this study are considered non-clinical because no visible disease presentation was recorded. However, 19% of the milk samples had a high SCC range of $\geq$ 400,000 cells/ml which suggests subclinical infection was present. Hunt *et al.,* (2011) reported no clinical signs in one human subject who had a bloom of *Streptococcus* representing 95% of the total bacterial relative abundance in the first milk sample, with a more even phylotype in the following two samples. This suggested a subclinical infection was present. Hence, elucidating the bacterial composition of individual milk samples and statistically modelling the composition against SCC could assist in deciphering the role of individual bacterial species in subclinical disease.

These previous studies on both cow and human milk provide an intriguing insight into bacterial community composition in milk and its variation over time and between individuals. However, most are cross-sectional studies with only one sample per subject. This limits the ability to further understanding of causality, that is how infection develops, which is of particular relevance when communities are highly personalised (Hunt *et al.*, 2011). Furthermore, none of those studies were conducted in sheep, so differences in community composition due to the differences in physiology and management of sheep might be expected.

This Chapter details the methodology for Illumina MiSeq analysis of 16S rRNA genes from a longitudinal study of sheep milk samples. The optimisation of the sequencing library preparation protocol, the stages in the subsequent data analysis pipeline and results for 5 sheep are presented and discussed.

5.2 Materials and methods

5.2.1 Illumina MiSeq library preparation protocol

Five libraries (up to 84 samples in each) were prepared for Illumina MiSeq paired-end sequencing using the two-step protocol summarised in Figure 5-1. All primer sequences used in library preparation are detailed in Table 5-1 and PCR programs in Table 5-2. All PCR reactions were prepared in a cabinet that was UV sterilised for 30 minutes prior to each use. Each PCR included a template PCR blank control. All reactions were carried out using Bioline MyTaq hot-start red mix detailed in Table 5-3 under standard conditions on an Eppendorf master cycler.

Briefly, the first PCR to attach required Illumina read sequences (these allow the MiSeq to read the indexes on the samples) was performed using primers 27F-rd1/338R-rd2. For some milk samples, no DNA and/or milk was present post PCR-DGGE analysis. Hence, to enable a fair comparison between samples and reduce the potential for PCR inhibition, 1µl of a 1:10 dilution of 27F/338RGC PCR product was used as the starting DNA template for the 27F-rd1/338R-rd2 PCR. Primer sequences, PCR program and Bioline MyTaq master mix components were detailed in Chapter 3 Section 3.2.2.

All 27F-rd1/338R-rd2 PCR products were purified using the Agencourt AMPure XP purification system (Beckman-Coulter) as per the manufacturer's instructions and 1µl of purified PCR product was used as template in the second N5/N7 PCR. The N5/N7 PCR conditions were based on those recommended in the Nextera XT DNA sample preparation kit (Illumina). An exemplar of the combination of N5/N7 indexes used per library is shown in Table 5-4. The samples submitted in each library are detailed in Appendix 5.

Post N5/N7 PCR, 10% of samples per library were selected for the P5/P7 check PCR using the random number function in Excel. 1µl of N5/N7 PCR product was used directly as DNA template in the P5/P7 check PCR. This PCR was to ensure that the indexes added in the N5/N7 PCR were in the correct orientation illustrated in Figure 5-3. Table 5-5 shows the combinations of primers tested in the P5/P7 check and the expected amplification result if indexes were in the correct orientation.

PCR products were then purified a second time using the Agencourt AMPure XP purification system (Beckman-Coulter) and normalised to 1-2 ng/µl using the SequalPrep kit (Invitrogen, UK) as per the manufacturer's instructions. Normalised samples for each library were pooled by adding 10µl of each sample to a 1.5ml microcentrifuge tube. Each pooled library was quantified using the Qubit (Invitrogen, UK) which gave a concentration reading of approximately 2nM. To concentrate each pooled library to the required 4nM, 20µl of each library was centrifuged at room temperature for 20 minutes on an Eppendorf vacuum concentrator. All five libraries were submitted to the Genomics facility at the School of Life Sciences, University of Warwick, for processing using the Illumina MiSeq machine.

Figure 5-1: Key steps in Illumina MiSeq library preparation of milk samples

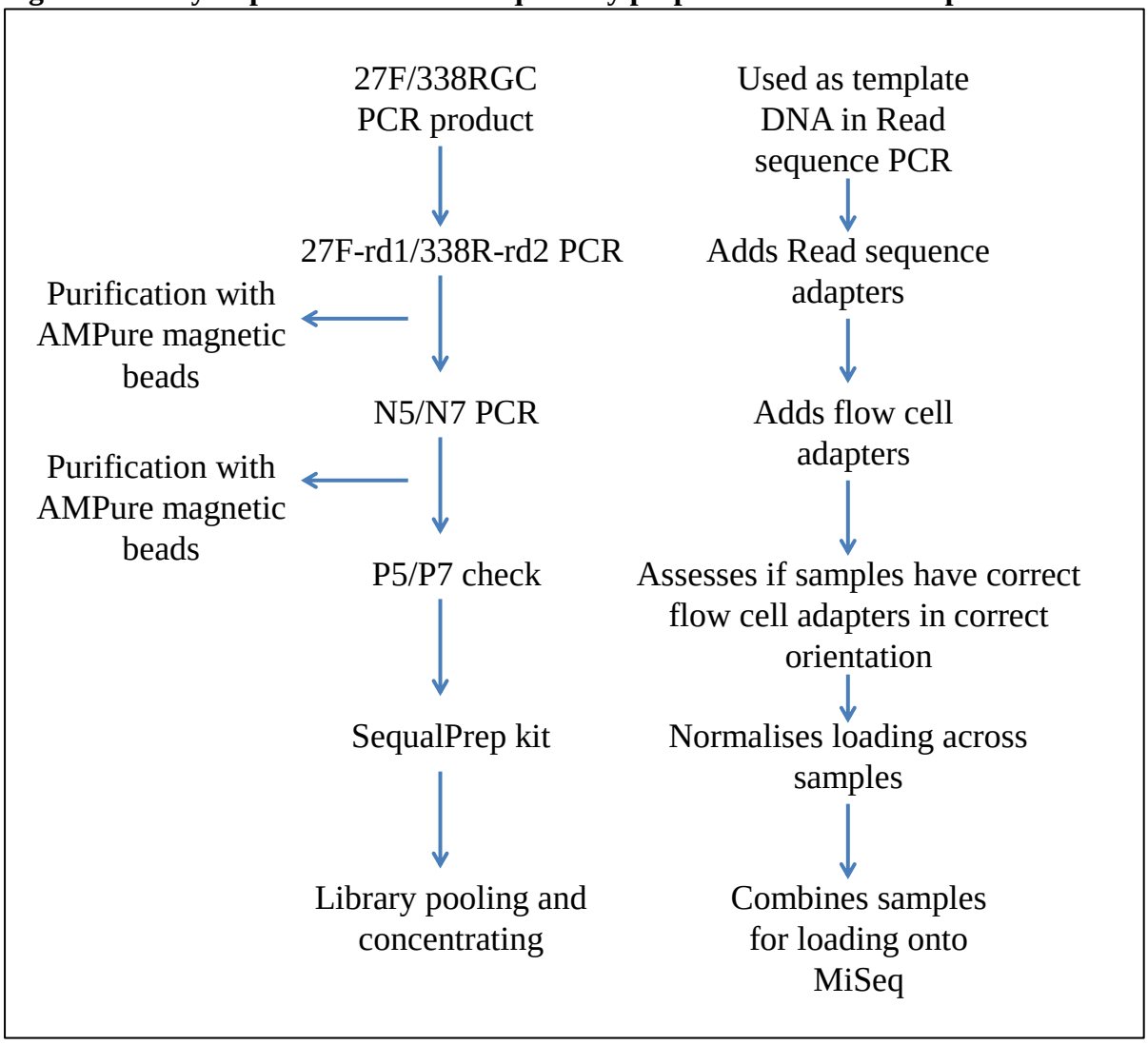


Table 5-1: PCR primers for Illumina MiSeq paired-end sequencing library preparation

Primer	Sequence 5'-3'	Annealing temp (^{o}C)	Product size (bp)
27F-rd1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGTTTGATCCTGGCTCAG		
338R-rd2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGCTGCCTCCCGTAGGAGT	61	406
N501-rd1	AATGATACGGCGACCACCGAGATCTACACGCGATCTATCGTCGGCAGCGTCAGATGT		
N502-rd1	AATGATACGGCGACCACCGAGATCTACACATAGAGAGTCGTCGGCAGCGTCAGATGT		
N503-rd1	AATGATACGGCGACCACCGAGATCTACACAGAGGATATCGTCGGCAGCGTCAGATGT		
N504-rd1	AATGATACGGCGACCACCGAGATCTACACTCTACTCTTCGTCGGCAGCGTCAGATGT		
N505-rd1	AATGATACGGCGACCACCGAGATCTACACCTCCTTACTCGTCGGCAGCGTCAGATGT		
N506-rd1	AATGATACGGCGACCACCGAGATCTACACTATGCAGTTCGTCGGCAGCGTCAGATGT		
N507-rd1	AATGATACGGCGACCACCGAGATCTACACTACTCCTTTCGTCGGCAGCGTCAGATGT		
N508-rd1	AATGATACGGCGACCACCGAGATCTACACAGGCTTAGTCGTCGGCAGCGTCAGATGT		
N701-rd2	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGGAGATGT		
N702-rd2	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGGAGATGT		
N703-rd2	CAAGCAGAAGACGGCATACGAGATAGGCAGAAGTCTCGTGGGCTCGGAGATGT		
N704-rd2	CAAGCAGAAGACGGCATACGAGATTCCTGAGCGTCTCGTGGGCTCGGAGATGT		
N705-rd2	CAAGCAGAAGACGGCATACGAGATGGACTCCTGTCTCGTGGGCTCGGAGATGT		
N706-rd2	CAAGCAGAAGACGGCATACGAGATTAGGCATGGTCTCGTGGGCTCGGAGATGT		
N707-rd2	CAAGCAGAAGACGGCATACGAGATCTCTCTACGTCTCGTGGGCTCGGAGATGT		
N708-rd2	CAAGCAGAAGACGGCATACGAGATCAGAGAGGGTCTCGTGGGCTCGGAGATGT		
N709-rd2	CAAGCAGAAGACGGCATACGAGATGCTACGCTGTCTCGTGGGCTCGGAGATGT		
N710-rd2	CAAGCAGAAGACGGCATACGAGATCGAGGCTGGTCTCGTGGGCTCGGAGATGT		
N711-rd2	CAAGCAGAAGACGGCATACGAGATAAGAGGCAGTCTCGTGGGCTCGGAGATGT		
N712-rd2	CAAGCAGAAGACGGCATACGAGATGTAGAGGAGTCTCGTGGGCTCGGAGATGT	55	516
P5 check	AATGATACGGCGACCACCGAGATCTACAC		
P7check	CAAGCAGAAGACGGCATACGAGAT	55	569

Table 5-2: PCR programs for Illumina MiSeq primers

Primers	PCR program
27F-rd1 338R-rd2	95°C for 1 min (hot-start), 94°C for 5 min, followed by 8 cycles of 94°C for 1 min, 61°C for 30 secs, 72°C for 2 min, followed by 2 min final extension at 72°C
N5-rd1 - N7-rd2	95°C for 1 min (hot-start), 72°C for 3 min, 95°C for 30 secs followed by 8 cycles of 95°C for 10 secs, 55°C for 30 secs, 72°C for 30 secs, followed by 5 min final extension at 72°C
P5/27F P5/338R P7/27F P7/338R	94°C for 5 min, followed by 10 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min, followed by 2 min final extension at 72°C

Table 5-3: Bioline MyTaq hot-start red master mix components

Master mix component	Working concentration	Reaction volume (50µl)
Primer forward	10µM	0.5
Primer reverse	10µM	0.5
MyTaq master mix	N/A	25
DNA	-	1
Water	-	23
Total (µl)	-	50

Figure 5-2: Structure of PCR amplicon after 2-step library preparation protocol

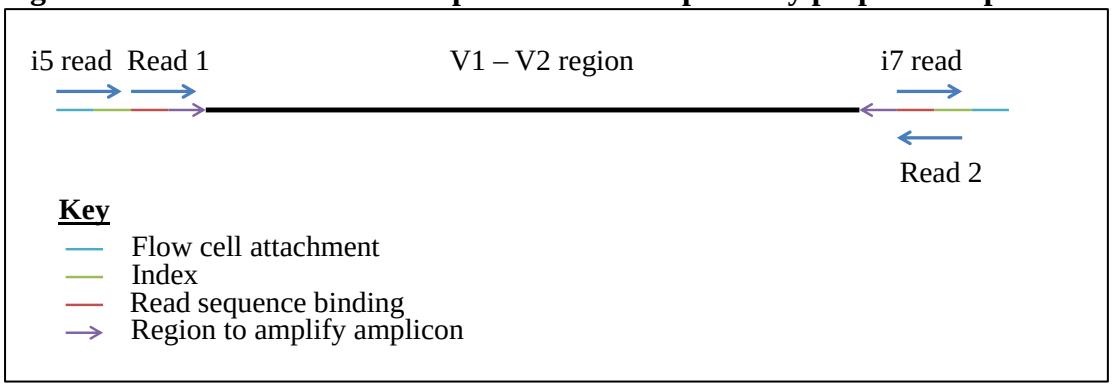


Figure 5-3: PCR amplicon for P5/P7 check PCR

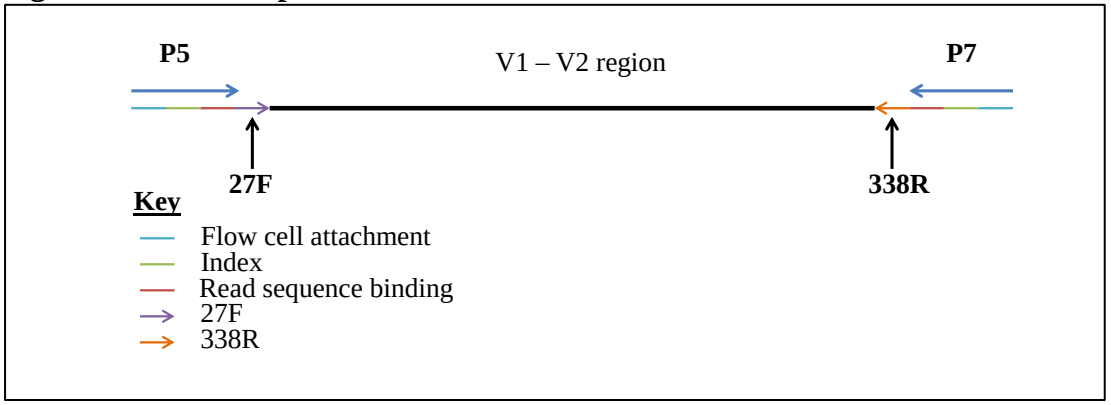


Table 5-4: Exemplar of N5/N7 PCR index combinations used per submitted library [25]

1			1	2	3	4	5	6	7	8	9	10	11	12
			N701	**N702**	**N703**	**N704**	**N705**	**N706**	**N707**	**N708**	**N709**	**N710**	**N711**	**N712**
	Sequence		**TCGC CTTA**	**CTAG TACG**	**TTCT GCCT**	**GCTC AGGA**	**AGGA GTCC**	**CATG CCTA**	**GTAG AGAG**	**CCTC TCTG**	**AGCG TAGC**	**CAGC CTCG**	**TGCC TCTT**	**TCCT CTAC**
B	**N502**	**ATAG AGAG**	A45 R1050	A46 R576	A29 L1101	A45 R1050 TR2	A47 L1721 RR1	A15 L27	A41 R266	A47 L1721 RR2	A20 L837	A35 L1277	A46 L289	A43 R736
C	**N503**	**AGAG GATA**	A26 R226	A45 R1712	A22 R1072	A25 R794	A45 R1050 TR1	A24 R696	A16 R1212	A40 R1266	A12 L743	A2 L819	A41 L1109	A20 R740
D	**N504**	**TCTA CTCT**	A12 R1296	A28 L1183	A23 L561	A35 L67	MC2	A28 R826	A162 L557	A26 L875	A16 R840	A26 R670	A23 R1706	A16 R318
E	**N505**	**CTCC TTAC**	A15 L1047	A43 L1035	A27 L663	A27 R664	A39 L1139	A43 L1197	A47 R1214	A47 L91	A22 R746	A35 R1150	A43 R1198	A21 L1159
F	**N506**	**TATG CAGT**	A49 R1714	-	A47 R726	A16 L1297	A23 L1043	A41 R882	A43 R704	A26 L457	A35 L331	A41 R672	A45 R1206	A29 L877
G	**N507**	**TACT CCTT**	A41 L671	A29 R878	A16 R716	A40 R790	A35 R1186	A25 L721	A15 L553	A47 L1065	A20 R1056	A21L3	A47 L1721	A27 L771
H	**N508**	**AGGC TTAG**	A26 R776	A34 R1020	A12 R844	A12 L1217	A25 L827	A43 R830	A23 L483	Model Community	A23 R856	A22 R264	A47 R92	Model Community

Table 5-5: Primers tested in P5/P7 check

Primer combination	Expected result
P5/27F	✗ No product
P7/338R	✓ ~ 400bp product
P7/27F	✓ ~ 400bp product
P7/338R	✗ No product

[25] In Table 5-4, the index sequence for each N5 and N7 primer is provided for library 1. The sample information is detailed: Sheep identification; mammary gland half; sample number. Model Community is the model community control which is discussed in further detail in Section 5.2.1.1. All sample locations were randomised across the 5 libraries using the random number function in Excel. A sample well marked '-' was empty.

5.2.1.1 Illumina MiSeq library model community control

Each library contained a model community. The model community consisted of 5 bacterial species commonly associated with intramammary infections in sheep; *Escherichia coli*, *Staphylococcus hyicus*, *Mannheimia haemolytica*, *Staphylococcus aureus* and *Streptococcus uberis*. DNA from each bacterial species was extracted from a pure culture using the DNeasy Blood and Tissue kit (Qiagen, UK). DNA concentration was measured using the NanoDrop spectrometer (Thermo Fisher Scientific, Loughborough, UK) and normalised to 25ng/µl. Each bacterial species underwent PCR with 27F/338R using the same conditions used for the sheep milk DNA samples detailed in Chapter 3 Section 3.2.2. PCR product for each bacterial species was purified using the QIAquick PCR purification kit (Qiagen, UK) then quantified again as shown in Table 5-6 and normalised to 25ng/µl. Normalised PCR product for each bacterial species was then combined in equal volumes to produce a model community sample to add to each of the 5 libraries for MiSeq analysis.

Table 5-6: DNA concentrations [ng/µl] and 260/280 ratios of bacterial species used to produce the model community positive control

Bacterial species	DNA concentration [ng/µ]	260/280 ratio
Staphylococcus aureus	167.6	1.91
Streptococcus uberis	159.7	1.90
Escherichia coli	160.5	1.95
Staphylococcus hyicus	158.0	2.06
Mannheimia haemolytica	170.8	2.06

5.2.2 Illumina MiSeq data analysis pipeline

Raw sequence data was edited using a combination of custom Perl and shell scripts and open-source software packages USEARCH (Edgar, 2010), UPARSE (Edgar, 2013) and QIIME (Caporaso *et al.*, 2010). Briefly, the forward and reverse read sequences for each sample were merged in USEARCH with 1 mismatch allowed and forward and reverse reads truncated at the first quality score <=3 to remove low quality tails. A custom Perl script re-labelled sequence headers in merged files which were then quality-filtered in USEARCH. In the quality-filtering process, sequence read length was analysed using the program Read Length Incremental Clustering (ReLIC). ReLIC determines whether sequences of specific lengths around the median read length lead to large changes in the number of OTUs detected (Williams and Purdy, submitted). This analysis selected $\leq$312bp as the minimum acceptable read length for this dataset. In addition, a maximum error rate of 0.065 was used to determine

whether sequences were of sufficient quality. Representative sequences for each sample were identified using dereplication scripts in USEARCH. Dereplicated files for individual samples were clustered into OTUs, re-labelled and chimera checked in UPARSE. This analysis (see Section 5.3.3.) gave a confidence limit for a minimum OTU size of 0.43% of total reads determined using analysis of the dereplicated model community across the 5 sample libraries. OTUs below this size cut-off were removed from each sample before reads were mapped back to each sample from the quality-filtered file using the UPARSE pipeline to produce OTU tables. To produce the initial analysis of 5 sheep shown in Section 5.3.4, an OTU table per sheep was produced by merging OTU tables for individual samples using scripts in QIIME.

5.3 Results

5.3.1 Optimisation of MiSeq library preparation protocol

5.3.1.1 Read sequence addition (27f-rd1/338R-rd2) PCR optimisation

A number of changes to the 27F-rd1/338R-rd2 PCR protocol (based on 27F/338R PCR) were made to determine the conditions that would achieve sufficient amplification to visualise via agarose gel electrophoresis for each sample.

The first optimisation was cycle number. The aim was to identify the minimum number of PCR cycles that would produce sufficient amplification to minimise PCR bias and/or chimera formation. Cycle number was tested in 5 cycle increments from 5-35 cycles. A cycle number $\geq$20 produced excessive PCR product and non-specific banding. Figure 5-4 shows the results of PCR cycle numbers of 5 to 20 in 5 cycle increments. It was determined from Figure 5-4 that 15 PCR cycles or more produced a non-specific band and 5 cycles produced insufficient amplification, making 10 cycles the number of choice. To determine if cycle number could be reduced below 10 and >5, Figure 5-5 shows the results of testing 8 cycles. As 8 cycles produced visible amplification with a good PCR yield, 8 cycles was chosen as the final cycle number.

Figure 5-4: 27F-rd1/338R-rd2 PCR results of testing 5-20 cycles [26]

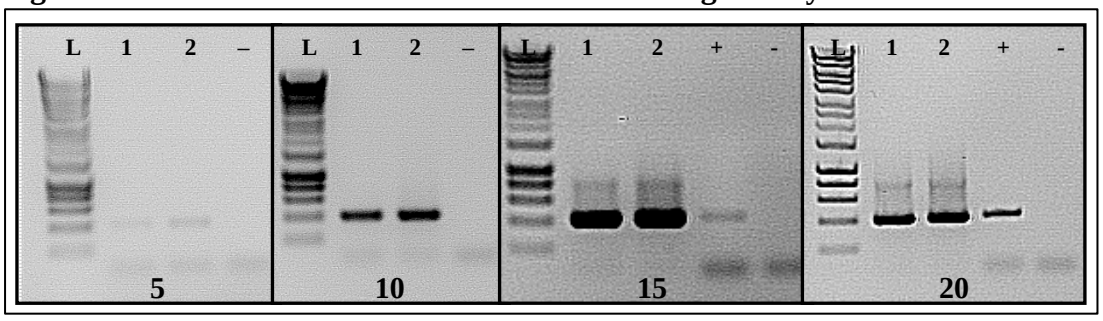
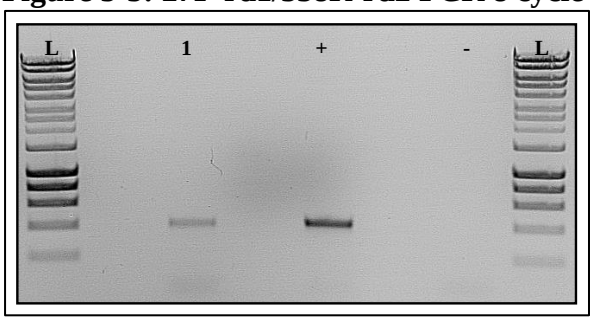


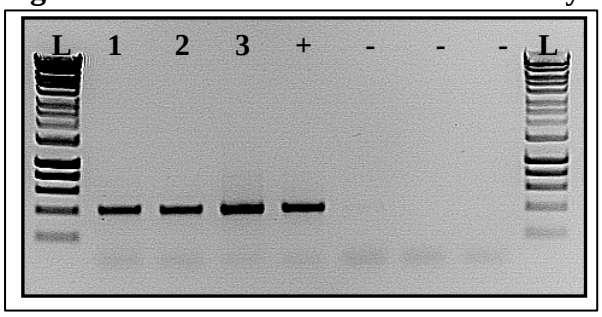
Figure 5-5: 27F-rd1/338R-rd2 PCR 8 cycle test results



[26] In both Figure 5-4 and Figure 5-5, the 'L' is Hyperladder 1kb (Bioline, UK). Numbers 1 and 2 are milk samples; '+' and '-' are the PCR positive and negative controls respectively.

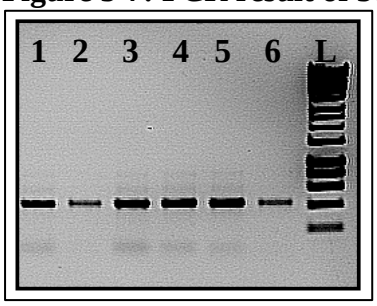
The second optimisation to the 27F-rd1/338R-rd2 PCR was to determine if a hot-start version of MyTaq would improve PCR yield. Figure 5-6 shows that the 27F-rd1/338R-rd2 at 8 cycles with hot-start Taq produces more intense PCR product in comparison to the standard MyTaq in Figure 5-5. Hence, the library PCR protocol changed to use the hot-start Taq.

Figure 5-6: 27F-rd1/338R-rd2 PCR at 8 cycles with hot-start Taq [27]



A temperature gradient from 56°C-66°C was completed using the model community positive control. The aim was to use the lowest annealing temperature to reduce preferential amplification whilst maintaining PCR specificity (Sipos *et al.*, 2007). Figure 5-7 shows the PCR amplification from 56°C (1) - 61°C (6). In Figure 5-7, the increase in annealing temperature from 59°C (lane 5) to 61°C (lane 6) resulted in the removal of non-specific bands. Therefore, the final annealing temperature used for this PCR in library preparation was 61°C.

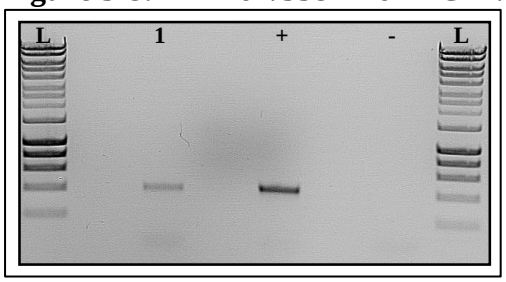
Figure 5-7: PCR result of 56°C- 61°C 27F-rd1/338R-rd2 temperature gradient



Long annealing times can increase the likelihood of non-specific amplification, with 30 seconds recommended in the literature (Yu and Pauls, 1992). Hence, decreasing the annealing time from 60 - 30 seconds was tested. Figure 5-8 shows no visual difference in PCR product intensity when annealing time is reduced compared to the same PCR in Figure 5-5, resulting in a 30 second annealing time being used in the final 27F-rd1/338R-rd2 protocol.

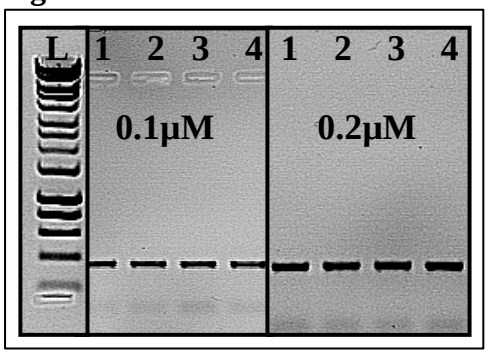
[27] 'L' is Hyperladder 1kb (Bioline, UK); 1-3 are milk samples; '+' is an *E. coli* positive control and '-' is the PCR negative control.

Figure 5-8: 27F-rd1/338R-rd2 PCR with 30 second annealing time at 8 PCR cycles



The final primer concentration used in the 27F-rd1/338R-rd2 PCR was tested at 0.2µM and 0.1µM. A reduced concentration was tested to minimise primer dimer formation without a decrease in PCR product yield. Figure 5-9 shows sufficient PCR yield with less prominent primer dimer at 0.1µM in comparison to 0.2µM. Hence, the lower primer concentration was used in the final protocol for MiSeq library preparation.

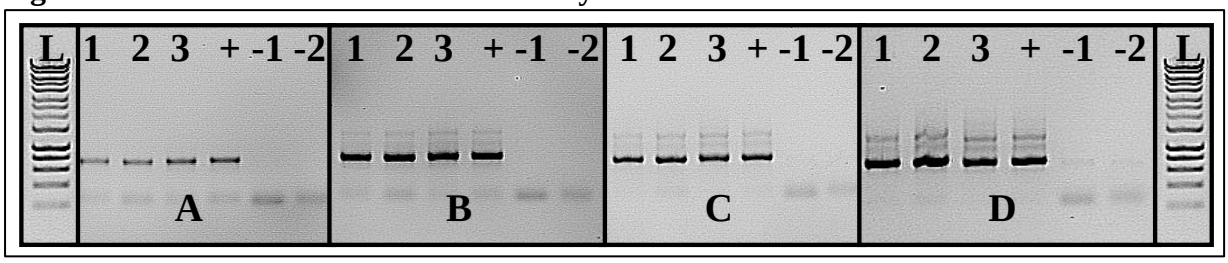
Figure 5-9: 27F-rd1/338R-rd2 PCR with 0.1µM and 0.2µM primer [28]



5.3.1.2 N5/N7 PCR optimisation

PCR cycle number was tested in the N5/N7 PCR using purified PCR product from the optimised 27F-rd1/338R-rd2 PCR. In Figure 5-10, the N5/N7 PCR was tested at 8 (A), 10 (B), 12 (C) and 15 (D) cycles. Only 8 cycles produced PCR product without additional non-specific banding. Hence, the final N5/N7 protocol was completed at 8 cycles.

Figure 5-10: N5/N7 PCR test of different cycle numbers [29]



[28] Numbers 1-4 are the mixed community positive control; 'L' is Hyperladder 1kb (Bioline, UK).

[29] 'L' is Hyperladder 1kb (Bioline, UK); 1-3 are sheep milk samples; '+' is the mixed community positive control; '-1' is the negative control from the mixed community PCR carried through the library preparation protocol and '-2' is the N5/N7 PCR negative control. 'A' is N5/N7 tested at 8 cycles, 'B' at 10 cycles, 'C' at 12 cycles and 'D' at 15 cycles, all with an 8 cycle 27F-rd1/338R-rd2 PCR beforehand.

5.3.2 Library preparation PCR results

Each library was successfully processed as described in Section 5.2.1. Figure 5-11 and Figure 5-12 show the PCR results from the 27F-rd1/338R-rd2 PCR for library 1.

Figure 5-13 and Figure 5-14 are the PCR results from the N5/N7 PCR for library 1. The PCR results for the other 4 libraries are in Appendix 6. The majority of the samples produced visible amplification. Those with weak amplification were repeated.

Figure 5-11: Results of 27F-rd1/338R-rd2 PCR for first 54 samples in library 1 [30]

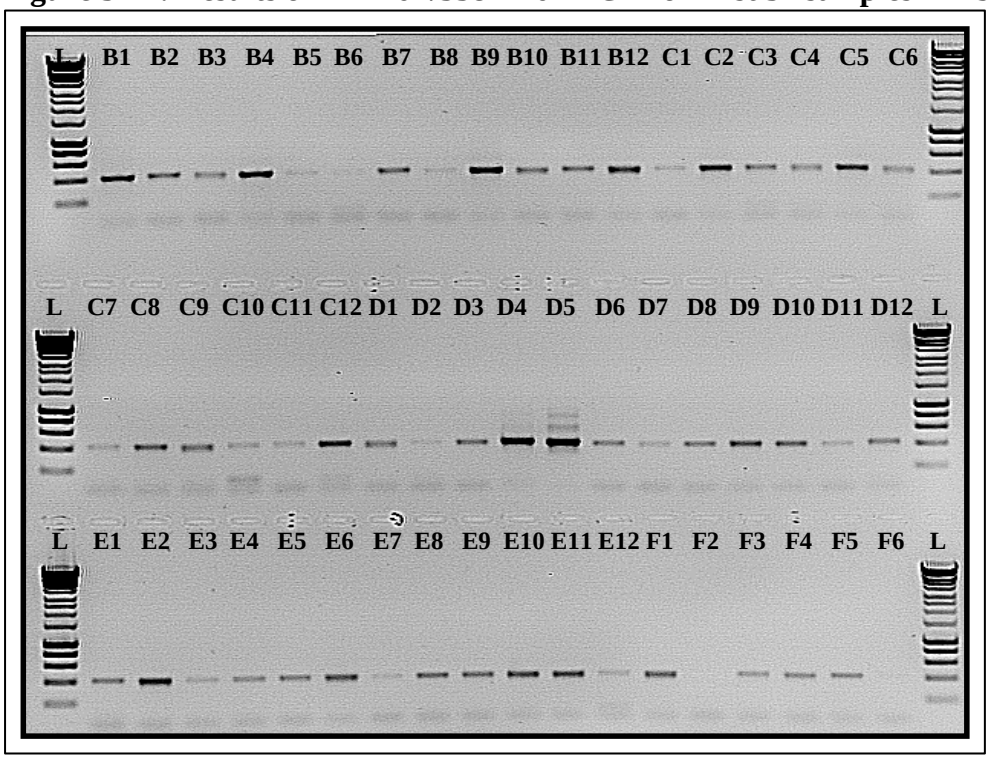
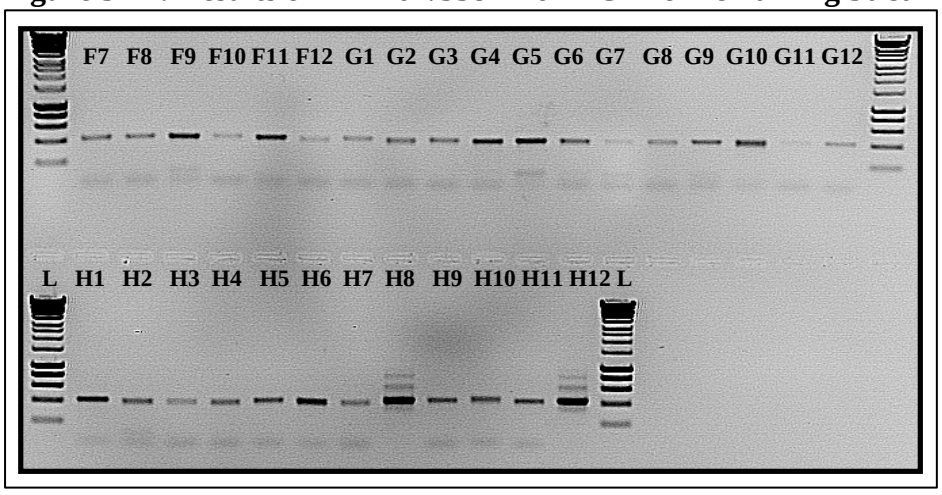


Figure 5-12: Results of 27F-rd1/338R-rd2 PCR for remaining 30 samples in library 1



[30] In Figure 5-11 and Figure 5-12, 'L' is Hyperladder 1kb (Bioline, UK); the number and letter for each sample correspond to the position of the sample on the 96-well plate for library shown in Table 5-4. Samples with weak amplification were repeated.

Figure 5-13: Results of N5/N7 PCR for first 54 samples in library 1 [31]

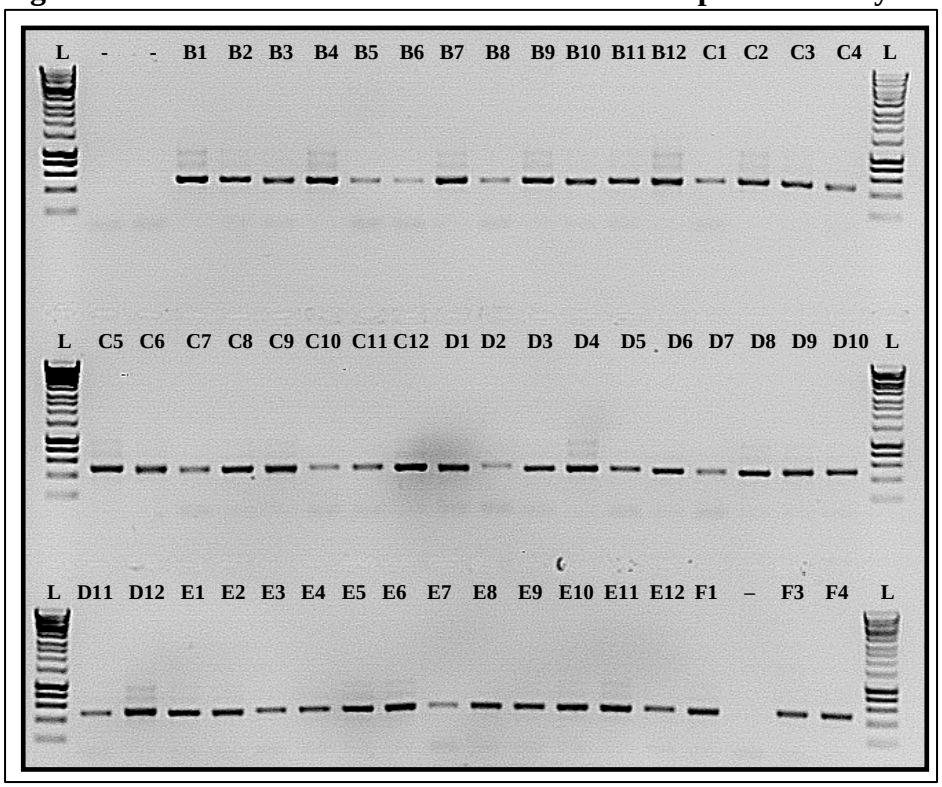
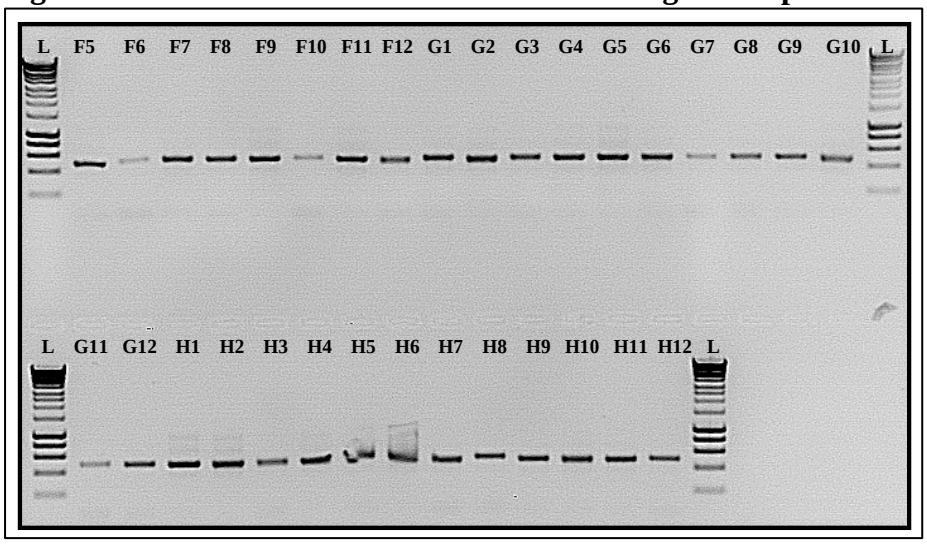


Figure 5-14: Results of N5/N7 PCR for remaining 30 samples in library 1



For the P5/P7 check PCR, 7 samples were picked using the random number function in Excel from each of the 5 libraries as listed in Table 5-7. The primer combinations tested in the P5/P7 check were detailed in Table 5-5. Figure 5-15 shows the results of testing the P5 and

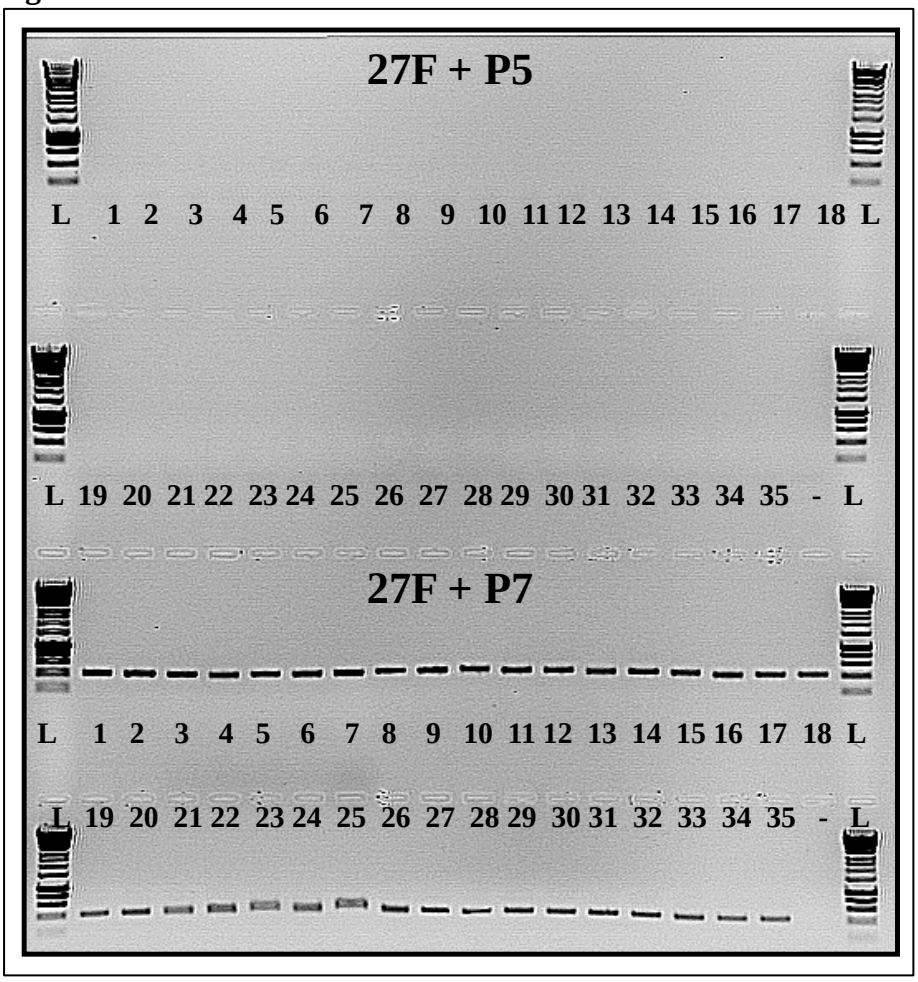
[31] In Figure 5-13, 'L' is Hyperladder 1kb (Bioline, UK); the letter and number for each sample correspond to their position on a 96-well template for library 1 as shown in Table 5-4. The 3 '-' symbols are negative controls carried through from the 27F-rd1/338R-rd2 PCR.

P7 indexes with 27F and Figure 5-16 shows the results of testing the P5 and P7 indexes with 338R. As expected, amplification was only seen for P7 + 27F and P5 + 338R.

Table 5-7: Samples chosen at random for each library for the P5/P7 check PCR [32]

Library	Sample
1	B1(1), C5(2), D4(3), E6(4), F11(5), G2(6), H2(7)
2	B3(8), C7(9), D1(10), E8(11), F8(12), G4(13), H6(14)
3	B9(15), C7(16), D3(17), E9(18), F8(19), G1(20), H5(21)
4	B7(22), C12(23), D3(24), E5(25), F6(26), G4(27), H3(28)
5	B5 (29), C8(30), D10(31), E10(32), F8(33), G2(34), H8(35)

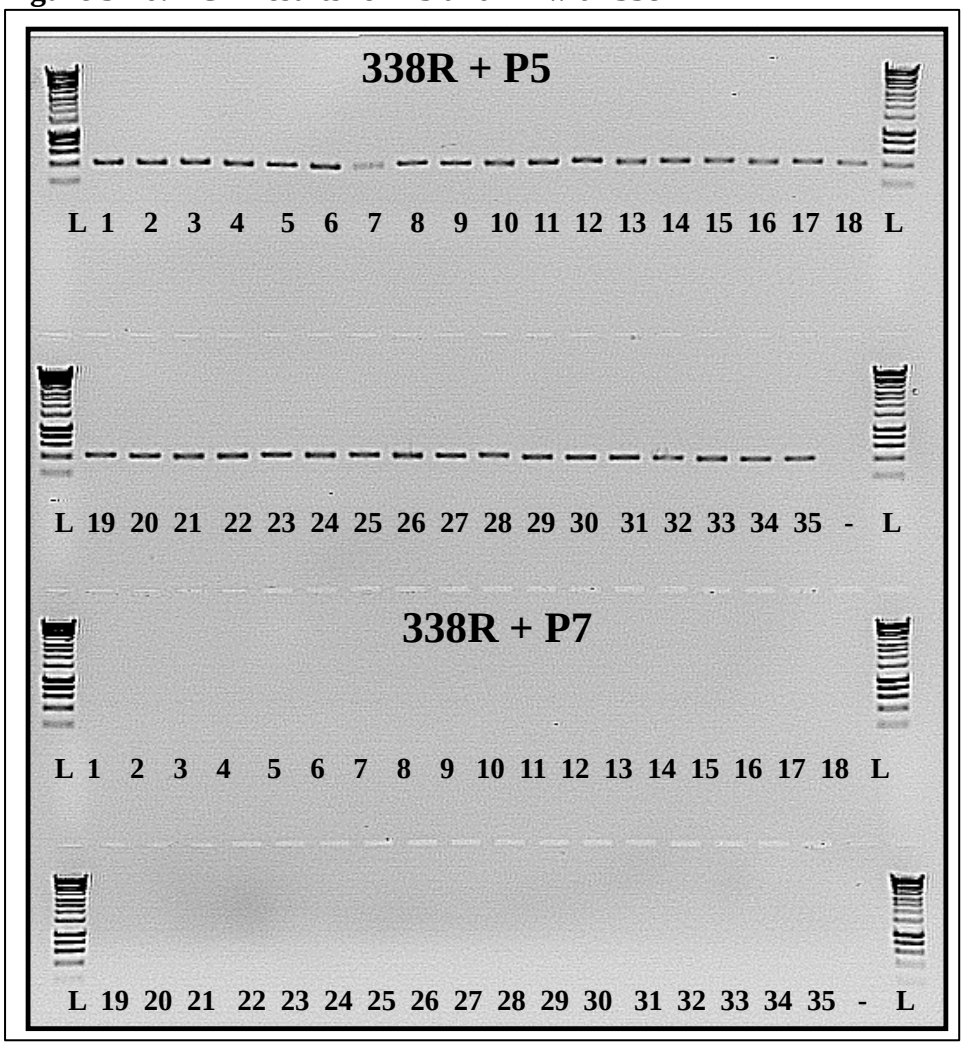
Figure 5-15: PCR results for P5 and P7 with 27F [33]



[32] Sample identifications correlate to their position on a 96-well template as shown in Table 5-4 for library 1 and in Appendix 5 for the other 5 libraries. The number in brackets corresponds to the number in both Figure 5-15 and Figure 5-16.

[33] The first two rows are for 27F + P5 and the second for 27F + P7. The numbers correspond to those in brackets in Table 5-7.

Figure 5-16: PCR results for P5 and P7 with 338R



5.3.3 Data analysis pipeline results

A total of 98,364,002 sequences were generated, with range of 42,734 - 266,618 sequences per sample and an average of 118,797 sequences per sample. The first step in the pipeline was the merging of the forward and reverse read for each sample. Table 5-8 summarises the range of raw reads per library. To determine the parameters to merge reads, different combinations of mismatch number and quality score truncation were tested on one sample (Table 5-9) and a model community control (Table 5-10).

Table 5-8: Minimum and maximum number of raw reads per library

Library	No. of reads min-max
1	63,257 - 225,809
2	80,798 - 266,618
3	42,734 - 132,512
4	55,725 - 192,464
5	50,741 - 214,223

Table 5-9: Merged reads parameter testing with milk sample L3 from sheep A2

Truncation score	No. of mismatches	Merged reads (%)
1	1	84.2
1	2	87.3
1	3	89.1
1	4	90.4
2	1	84.2
2	2	87.3
2	3	89.2
2	4	90.5
3	1	84.2
3	2	87.3
3	3	89.2
3	4	90.5
4	1	84.2
4	2	87.3
4	3	89.2
4	4	90.5

Table 5-10: Merged reads parameter testing with model community 1 from library 1

Truncation score	No. of mismatches	Merged reads (%)
1	1	83.3
1	2	87.1
1	3	89.4
1	4	91
2	1	83.3
2	2	87.2
2	3	89.4
2	4	91
3	1	83.3
3	2	87.2
3	3	89.4
3	4	91
4	1	83.3
4	2	87.2
4	3	89.4
4	4	91

Both Table 5-9 and Table 5-10 show that as the truncation score increases, the percentage of merged reads remain the same. USEARCH recommends a truncation score of at least 2 for paired-end reads; because there was no difference in the percentage of merged reads between scores 3 and 4, a score of 3 was chosen. When the number of mismatches was varied from 1-4, the percentage of reads merging varied up to 6.3% in Table 5-9 and 7.7% in Table 5-10. A stringent parameter of allowing 1 mismatch still resulted in 80-85% of reads merging. Consequently, 1 mismatch was allowed when merging reads. After merging, the number of sequences per sample ranged from 34,562 - 220,943, with an average of 98,794 sequences per sample. Table 5-11 summarises the number of reads merged per library.

Table 5-11: Minimum and maximum number of reads merged per library

Library	No. of reads merged min-max
1	52528 - 174140
2	64895 - 220943
3	34562 - 101167
4	46735 -169042
5	43513 -182743

After the sequence headers in each sample were re-labelled using a custom Perl script, the data were quality filtered based on the sequence length and maximum error rate. Figure 5-17 and Figure 5-18 show the sequence length distribution for a milk sample and model community.

Figure 5-17: Read length distribution for sheep A2 sample L3

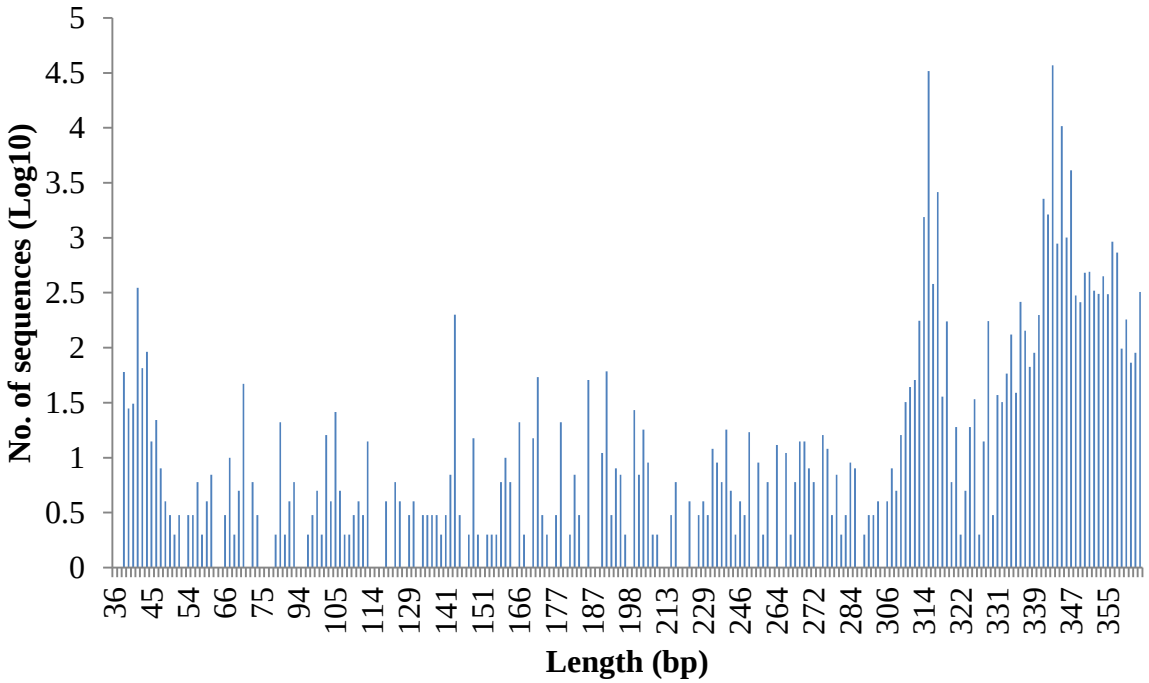
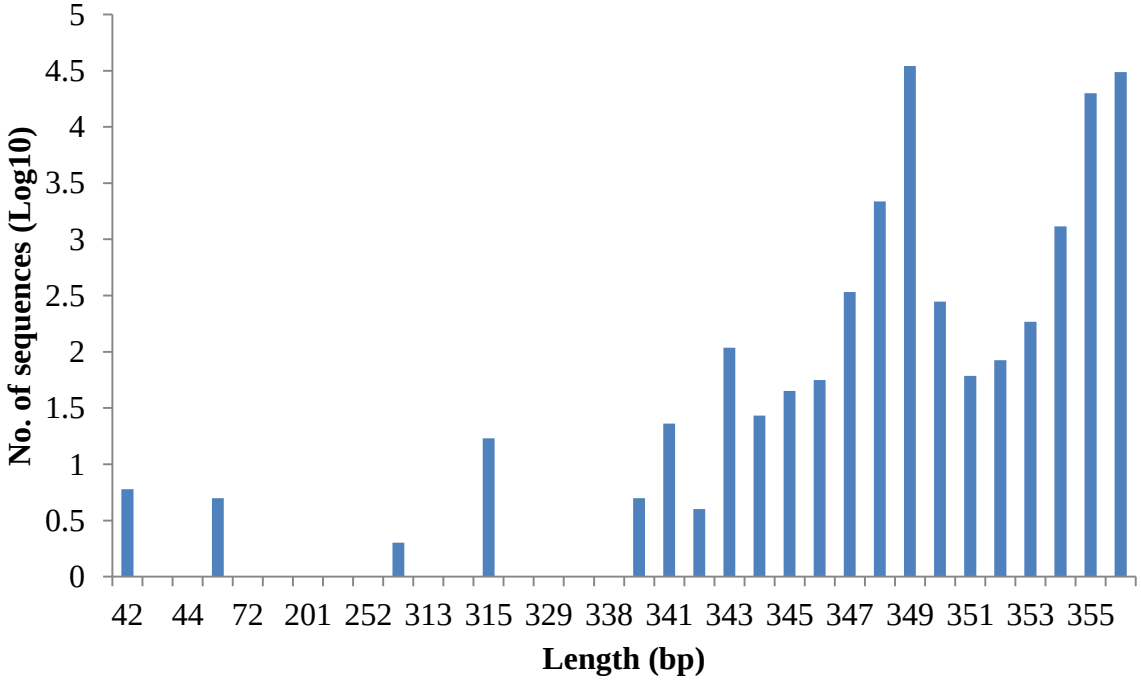


Figure 5-18: Read length distribution for model community 1



As shown by comparison of Figure 5-17 with Figure 5-18, the range in sequence length was greater in the milk sample in comparison to the model community positive control; a trend observed in all the milk sample data. Despite this, both show that the majority of sequences are greater than 300bp in length; with two peaks in sequence length at approximately 315bp and 340bp (median sequence length was 345bp). A Read Length Incremental Clustering (ReLIC) analysis (Williams and Purdy, submitted) of both samples determined a minimum

sequence length of 312bp represented biologically relevant diversity (data not shown). Therefore, a minimum sequence length of 312bp was selected for quality filtering.

The second quality filtering parameter was the maximum error rate. To determine the most stringent error rate, the sample with the smallest number of reads per library for libraries 1, 2, 4 and 5 were selected. A total of 6 different error rates were selected. The data for each sample was quality filtered with a minimum sequence length of 312bp and each of the error rates. The data were then dereplicated and clustered into OTUs. As shown in Table 5-12, there was a large decrease in the number of OTUs and sequences when the maximum error rate was reduced from 0.06 to 0.05, suggesting that the error rate could not be reduced below 0.06. To investigate further, the model community positive control from each plate was also tested at the six different error rates as shown in Table 5-13.

Table 5-12: Testing a milk sample per library at six maximum error rates

Library	Sample ID	Error rate	No. of OTUs	No. of sequences
1	A2-5-R794	0.1	339	40691
		0.09	332	40680
		0.08	321	40095
		0.07	299	38654
		0.06	217	34133
		0.05	19	13211
2	A2-6-R1102	0.1	88	47408
		0.09	86	46974
		0.08	85	46044
		0.07	83	43610
		0.06	67	36448
		0.05	18	12155
4	A46-1-R90	0.1	121	32800
		0.09	119	32385
		0.08	111	31503
		0.07	89	29208
		0.06	49	22434
		0.05	9	1903
5	A44-2-R50	0.1	1033	28916
		0.09	1027	28588
		0.08	1007	27734
		0.07	970	25686
		0.06	780	18671
		0.05	27	478

Table 5-13: Testing the model community from each library at 6 error rates

Library	Sample ID	Error rate	No. of OTUs	No. of sequences
1	Model community 1	0.1	13	64966
		0.09	13	64148
		0.08	12	62487
		0.07	12	58246
		0.06	10	43464
		0.05	0	6
2	Model community 1	0.1	15	46470
		0.09	15	45693
		0.08	14	43964
		0.07	12	39817
		0.06	11	27344
		0.05	0	13
4	Model community 1	0.1	15	72397
		0.09	14	71187
		0.08	14	68329
		0.07	13	61753
		0.06	12	41292
		0.05	1	10
5	Model community 1	0.1	17	54297
		0.09	16	53778
		0.08	16	52644
		0.07	16	49912
		0.06	13	38921
		0.05	3	25

Table 5-13 also shows a large decrease in the number of OTUs and reads when the error rate is reduced from 0.06 to 0.05. Hence, a conservative but stringent error rate of 0.065 was selected. Table 5-14 shows the range in read size for each of the 5 libraries post quality filtering. There were 26,279,490 sequences in total post quality filtering, with a minimum of 14,857, a maximum of 149,091 and an average of 63,477 sequences per sample.

Table 5-14: Range in number of reads per library post quality filtering

Library	No. of reads min-max
1	37098 - 122952
2	31155 - 149091
3	14857 - 62720
4	26623 - 117286
5	23409 - 121668

The composition of the model community was known as detailed in Section 5.2.1.1. Hence, the clustering of data into OTUs for the model community was used to assess the effectiveness of the quality filtering process. When a model community from each library

was quality filtered based on the 312bp sequence length cut-off and 0.065 error rate, between 9 and 14 OTUs were identified as shown in Table 5-15. This is greater than the 5 OTUs expected (5 bacterial species in model community).

Table 5-15: OTU identity and distribution across model community 1 for each library after remapping all reads onto the detected OTUs [34]

		MC1 L1	MC1 L2	MC1 L3	MC1 L4	MC1 L5	Total per OTU	% of total per OTU
OTU	**Blast Result**							
1	***Streptococcus uberis***	**12160**	**8695**	**6255**	**12219**	**10671**	**50000**	**24.11**
2	***Mannheimia haemolytica***	**9396**	**6560**	**4774**	**9068**	**8089**	**37887**	**18.27**
3	***Staphylococcus hyicus***	**9057**	**5524**	**4588**	**8054**	**7783**	**27223**	**13.13**
4	***Staphylococcus aureus***	**9401**	**6780**	**3022**	**9763**	**8462**	**37428**	**18.05**
5	***Escherichia coli***	**4223**	**2549**	**5386**	**4747**	**3654**	**20559**	**9.91**
6	*Escherichia coli/ Pseudomonas panacis*	8961	78	1525	68	29	10661	5.14
7	*Escherichia coli/ Pseudomonas panacis*	2	4	10	8275	20	8311	4.01
8	*Escherichia coli/ Pseudomonas panacis*	13	4725	0	5	6310	11053	5.33
9	*Escherichia coli/ Lacibacter spp./ Rhizobium spp.*	5	661	0	7	174	847	0.41
10	*Escherichia coli/ Limnobacter spp./ Burkholderia spp.*	2	2	0	1774	3	1781	0.86
11	*Escherichia coli/ Limnobacter spp./ Burkholderia spp.*	85	8	0	369	2	464	0.22
12	*Escherichia coli/* Rumen bacterium	0	0	0	0	2	2	0.00
13	*Alpha proteobacterium*	0	0	0	0	4	4	0.00
14	*Achromobacter xylosoxidans/ Escherichia coli*	0	0	0	0	1163	1163	0.56
Total	**-**	**53305**	**35586**	**25560**	**54349**	**46366**	**489913**	**-**

[34] 'MC1' refers to model community 1; 'L1' - 'L5' refers to library 1-5. The expected OTUs based on the model community composition are in bold.

Table 5-15 shows the 5 expected OTUs account for approximately 83% of the total reads. However, there were an additional 9 OTUs produced, many of which had the same inconclusive BLAST result suggesting the sequences in these OTUs could be chimeric sequences and/or PCR artefacts. This highlighted that the quality filtering of milk samples would also result in additional OTUs that did not represent biologically relevant diversity. The model community samples were therefore used to determine the OTU size that represented an unreliable cluster before remapping all sequences back onto the OTUs (Table 5-16).

Table 5-16: OTU sizes for model community 1 for all five libraries (dereplicated data) [35]

MC ID	No. of OTUs	No. of inputs	OTU size distribution	% of largest unreliable cluster	Chimeras	No. of reads with no match
MC1_P1	11	6847	3950, 3089, 2207, 1946, 1117, **4, 2(5)**	0.06	0	276
MC1_P2	11	4182	2843, 2116, 1394, 1371, 745, **18, 3(2), 2(3)**	**0.43**	0	167
MC1_P3	7	2751	1946, 1432, 893, **707, 528, 2(2)**	0.07	0	103
MC1_P4	13	5812	4071, 3099, 2073, 2056, 1384, **18, 4, 2(7)**	0.31	1	348
MC1_P5	14	6544	3169, 2575, 1744, 1621, 1016, **6, 4(2), 3, 2(5)**	0.09	1	279

From Table 5-16, the maximum OTU size that represented the first unreliable OTU was 0.43%. Subsequently, every milk sample MiSeq dataset was dereplicated and individually clustered into OTUs. Table 5-17 shows the range of sequence numbers for each of the 5 libraries post dereplication. There were 13,558,798 sequences in total, with a minimum of 7,546, a maximum of 80,972 and an average of 32,751 sequences per sample.

OTUs were chimera-checked and those that were equal to or lower than the determined confidence size of 0.43% of total reads were removed for each sample. Reads from the quality filtered file for each sample were then mapped back to the OTU file with all unreliable OTUs removed. Table 5-18 summarises the number of reads that were mapped back to filtered OTU files. Pre-OTU filtering, the minimum number of OTUs was 30, the

[35] 'MC ID' is model community identifier; 'No. of inputs' is the number of input sequences in the dereplicated file used to complete the OTU clustering; in 'OTU size distribution' the size of each OTU is defined; the number in brackets after the size refers to the number of OTUs that were of that size if there were multiple OTUs of the same size; the '% of largest unreliable cluster' was determined by dividing the size of the largest OTU that did not represent 1 of the 5 known bacterial species in the model community by the number of input sequences x100; the 'No, of reads with no match' is the number of sequences that did not map back to any of the OTUs post-filtering.

maximum 3,337 with an average of 728 OTUs per sample. After OTU filtering and mapping of reads, the minimum number of OTUs was 1, with a maximum of 22 and an average of 6 OTUs per sample. Table 5-19 shows the results of the OTU filtering process for parity 1 sheep A20 as an exemplar.

Table 5-17: Range in number of reads per library post dereplication

Library	No. of reads min-max
1	17023 - 60877
2	15244 - 80972
3	7546 - 31793
4	13849 - 51316
5	12949 - 66512

Table 5-18: Summary of mapping reads back to confident OTUs

Summary statistic	Mapped reads	Non-mapped reads
Minimum	3574	312
Maximum	126144	75178
Average	48216	15226
Total	20009726	6318880

Table 5-19: Sheep A20 results of OTU filtering process [36]

Sample	No. of inputs	No. of OTUs pre-filter	Chimeras	OTU size confidence	No. of OTUs post-filter	No. of reads mapped	No. of reads not mapped
1R	42837	2331	60	184	7	46618	27633
2L	33275	1370	33	143	8	53764	13869
2R	24912	1376	37	107	8	32398	10803
3L	27291	1278	36	117	5	42439	10997
3R	40153	1337	49	173	6	72339	10763
4L	13054	460	19	56	5	19330	2502
4R	59567	2181	69	256	5	90106	27847
5L	48613	1319	42	209	5	85897	15401
5R	30976	119	4	133	5	65713	3380
6L	44860	956	19	193	5	81959	11583
6R	19304	220	5	83	6	35253	3165
7L	30194	152	12	130	5	63122	3566
7R	21102	623	19	91	5	39203	4976
8L	37454	443	20	161	5	73494	4039
8L1	24696	415	23	106	5	43849	4072
8L2	24338	388	14	105	5	45656	6560

[36] Sample ID is defined by week 1-8 and 'L' or 'R' for left or right half; '8L1' and 8L2' are technical replicates of '8L'; 'No. of inputs' refers to the sequences in the dereplicated file used to cluster each sample into OTUs. Pre-filtered OTUs include singletons. Chimeras were detected post-OTU clustering; 'No. reads mapped' and 'No. reads not mapped' refer to the number of sequences per sample that did and did not map back to an edited OTU file.

Table 5-19 shows the large decrease in the number of OTUs as a result of the filtering process. Before OTU filtering, the minimum number of OTUs for a sheep A20 milk sample was 119, with a maximum of 2331 and an average of 935. This decreased to a minimum of 5 OTUs, a maximum of 8 and an average of 6 OTUs per sheep A20 milk sample. Table 5-19 also shows that despite this, the number of reads mapped back to an OTU is still greater than the number of reads that do not map, although in comparison to Figure 5-17, a much larger number of reads do not map back to filtered OTUs.

OTU sizes pre-filtering were investigated to visualise OTU confidence cut-off points.

Figure 5-19: OTU size distribution for model community from library 1

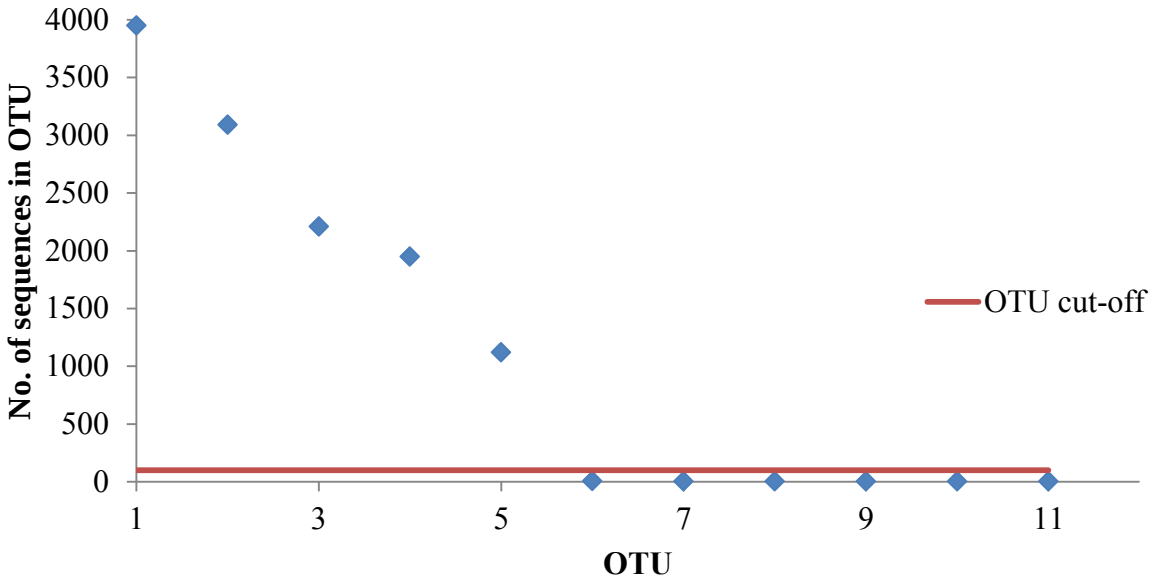
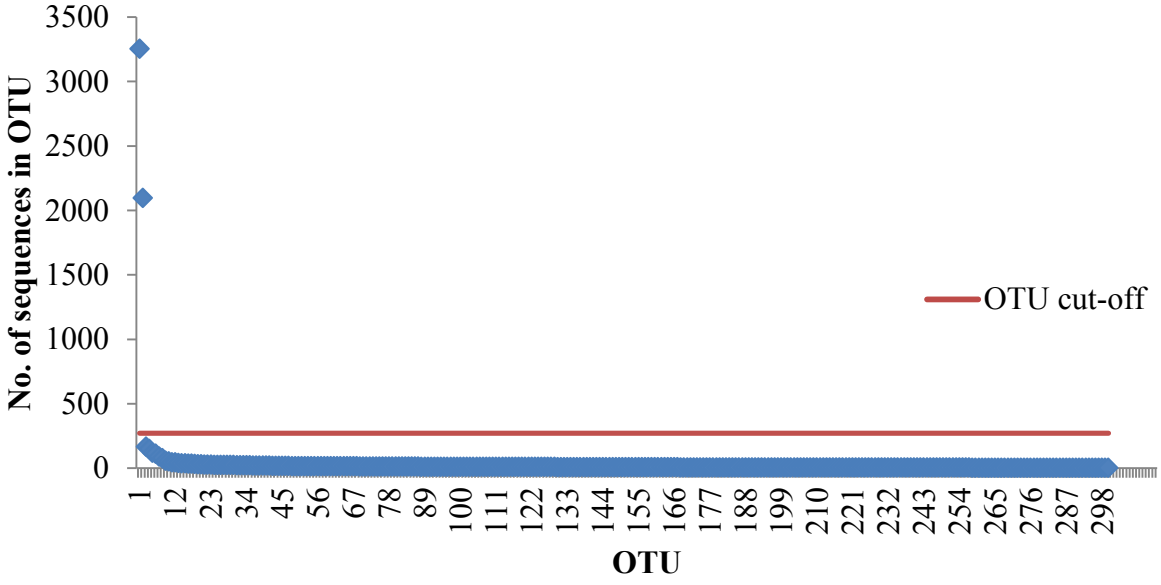


Figure 5-20: OTU size distribution for sheep A21 sample L1



In both Figure 5-19 and Figure 5-20, each OTU (x-axis) and the number of sequences assigned to that OTU (y-axis) were plotted. For the model community in Figure 5-19, there were 11 OTUs, with a large decline in OTU size between OTUs 5 and 6. The OTU cut-off size for the model community based on the 0.43% OTU confidence size was 86 (shown in Figure 5-19), resulting in only OTUs 1-5 only proceeding through the OTU filtering process. As 5 bacterial species were present in the model community, this indicated that the OTU filtering process produces an accurate representation of community composition. For the milk sample L1 in Figure 5-20, OTU 1 and 2 contained 3,253 and 2,095 sequences respectively, followed by a large decline in OTU size to 166 sequences in OTU 3. The OTU confidence size of 272 (shown in Figure 5-20) used to filter sample L1 meant only OTU 1 and 2 were retained. The sharp decline in OTU size in both examples below the OTU confidence size cut-off provides evidence to support the use of this approach for data filtering.

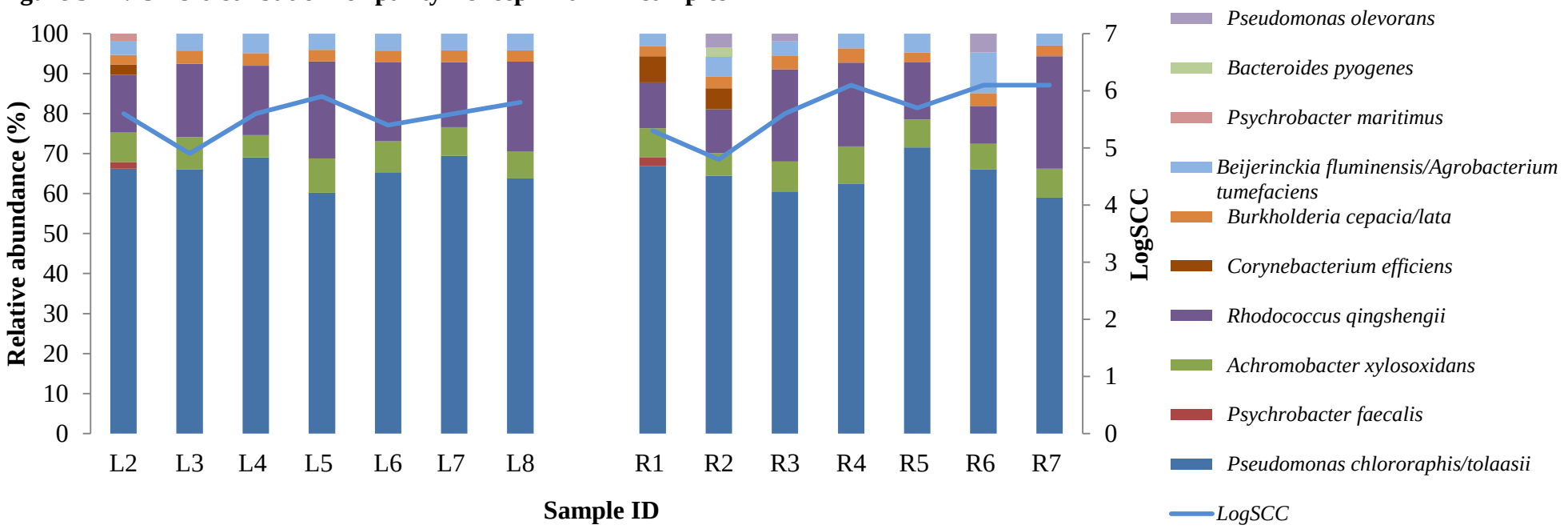
5.3.4 MiSeq results for 5 sheep

To illustrate the potential in the OTU filtered sequencing data set and to investigate the study hypotheses, the OTU identities and distributions were analysed for 5 sheep; one of each parity group.

Figure 5-21 - Figure 5-30 show the sequencing results for 5 sheep; (in increasing parity order); A20, A41, A25, A40 and A23. Each OTU distribution Figure is shown with the corresponding DGGE image for reference. Figure 5-31 and Figure 5-32 are technical replicates for samples L8 and L2 from sheep A20 and A23 respectively.

For sheep A20, 10 OTUs were identified, with *Pseudomonas* spp. dominating every milk sample. For parity 2 sheep A41, 34 OTUs were identified, with shifts in relative abundance of bacterial species from week to week within both halves, with *Corynebacterium efficiens* the most abundant in the majority of the milk samples. For parity 3 sheep A25, 10 OTUs were identified, with *Pseudomonas* spp. most abundant and a similar distribution of bacterial species across the weeks and mammary gland halves. Parity 4 sheep A40 had 6 OTUs with *Pseudomonas* spp. most abundant. For parity 10 sheep A23, 9 OTUs were identified, with *Pseudomonas* spp. and *Rhodococcus qingshengii* most abundant. The technical replicates in Figure 5-31 and Figure 5-32 show good consistency between replicates, with both sets of samples consisting of the same OTUs in similar abundances.

Figure 5-21: OTU distribution for parity 1 sheep A20 milk samples



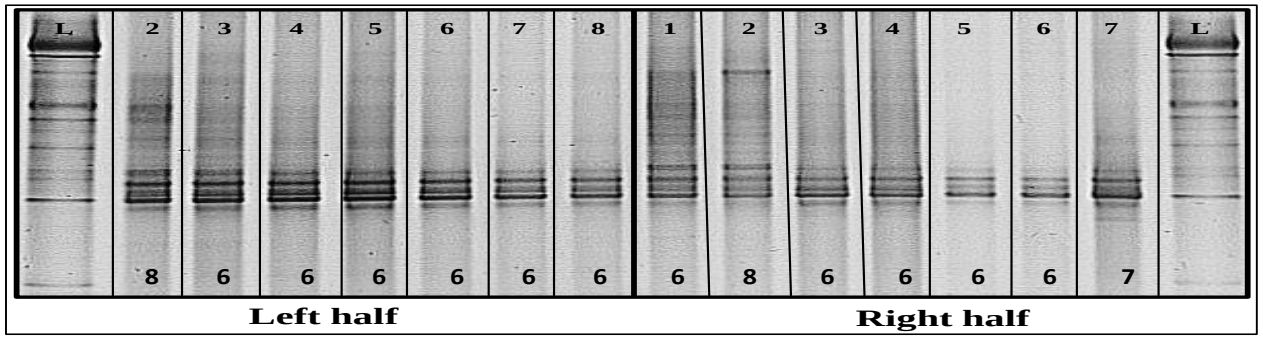
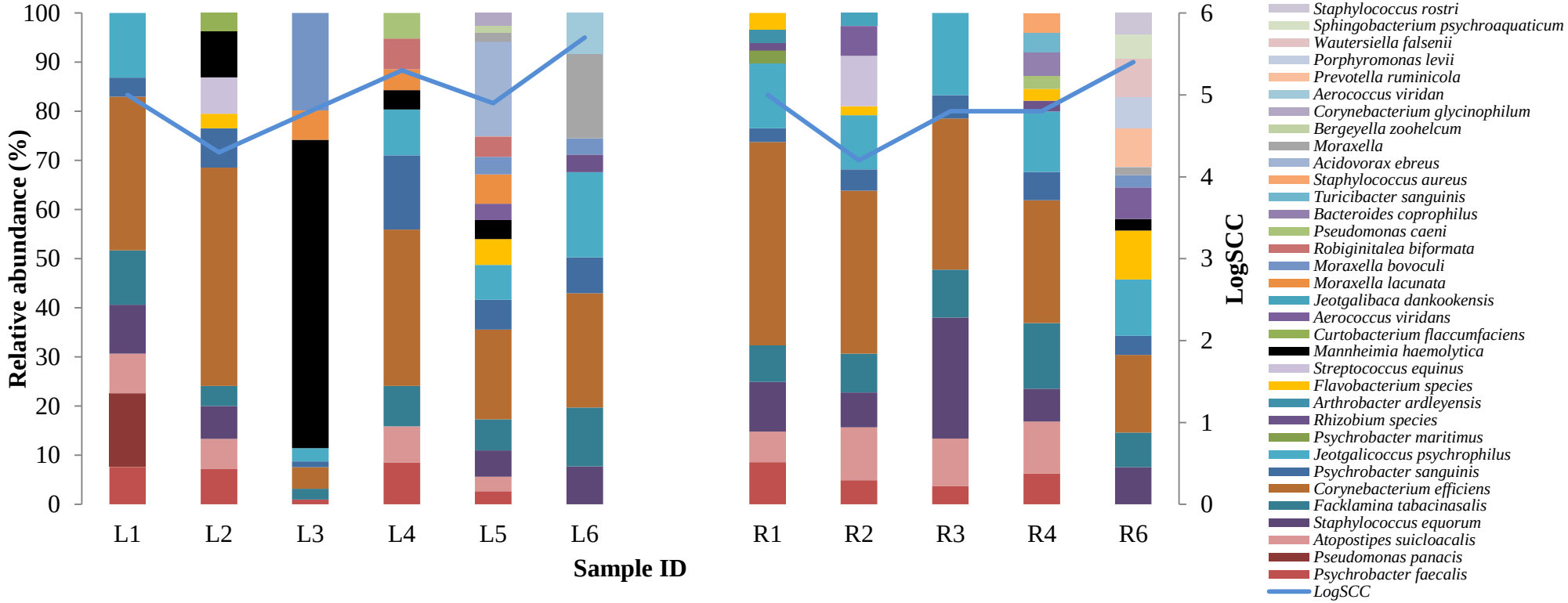


Figure 5-22: DGGE image of parity 1 sheep A20 [37]

[37] For all DGGE images in this section; 'L' is Hyperladder 1kb (Bioline, UK); the number at the top is the week in lactation and at the bottom is DGGE band count.

Figure 5-23: OTU distribution for parity 2 sheep A41 milk samples



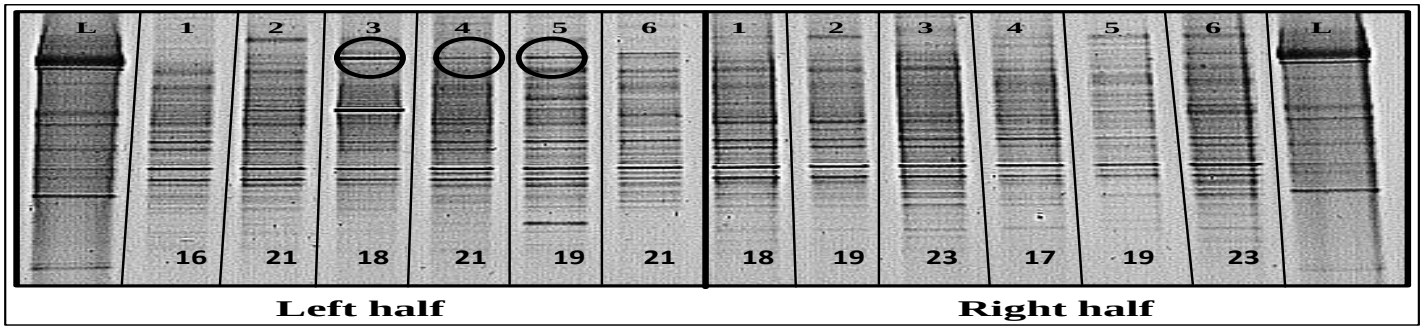
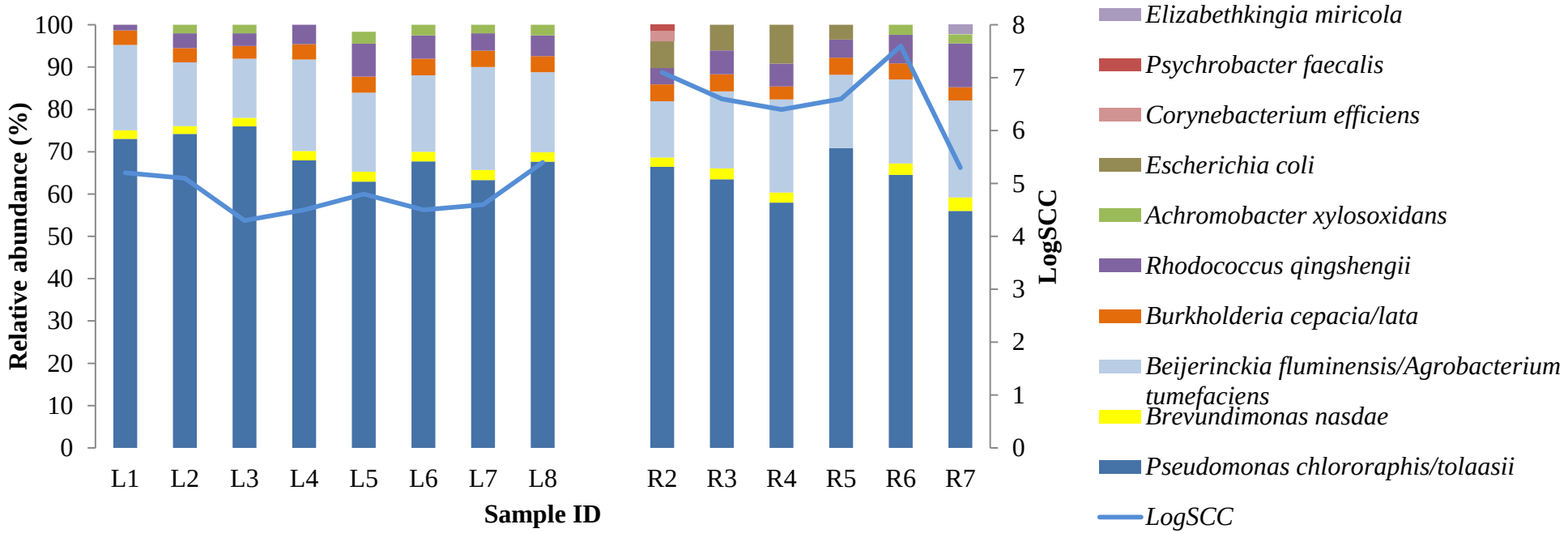


Figure 5-24: DGGE image for parity 2 sheep A41

Figure 5-25: OTU distribution for parity 3 sheep A25 milk samples



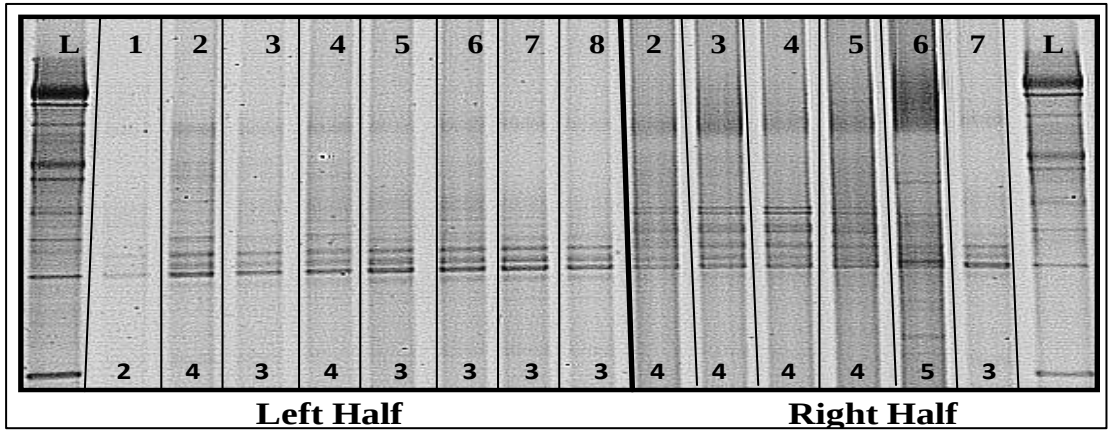
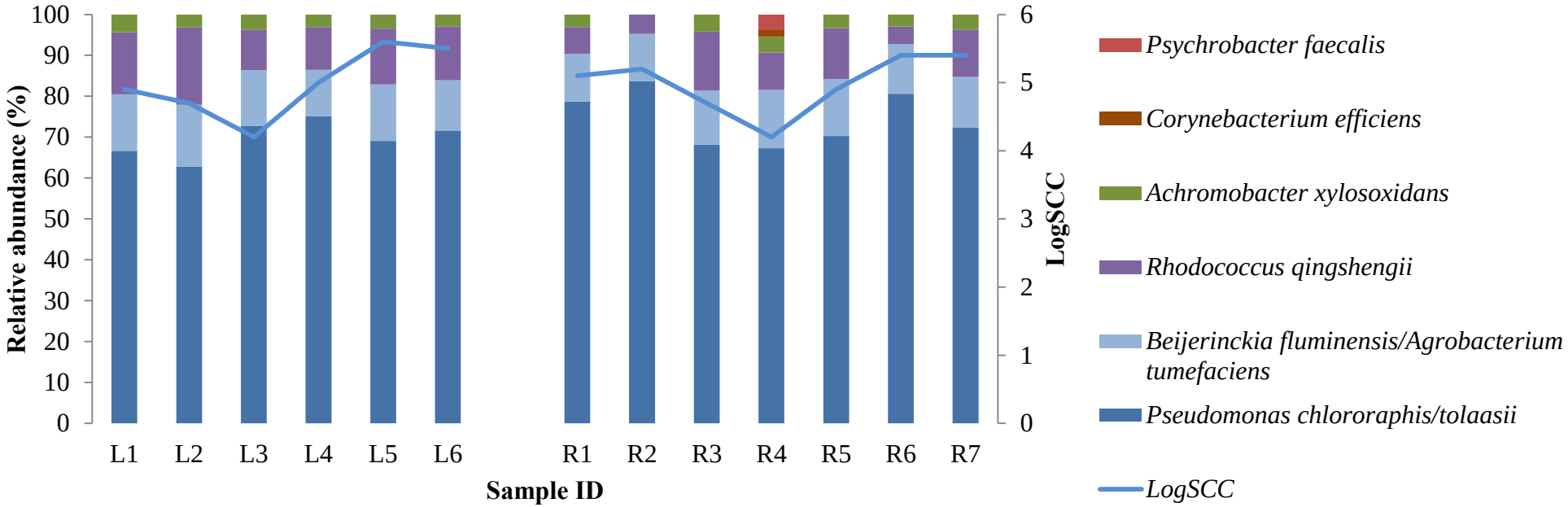


Figure 5-26: DGGE image for parity 3 sheep A25

Figure 5-27: OTU distribution for parity 4 sheep A40 milk samples



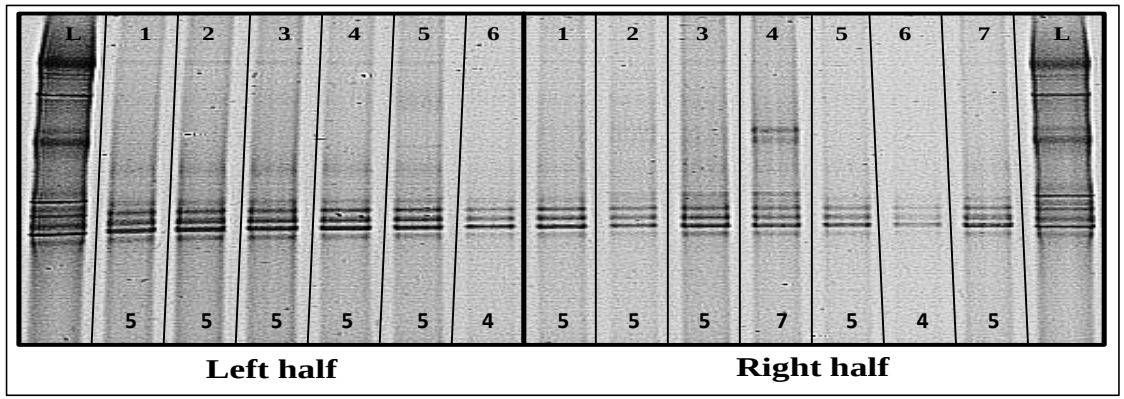
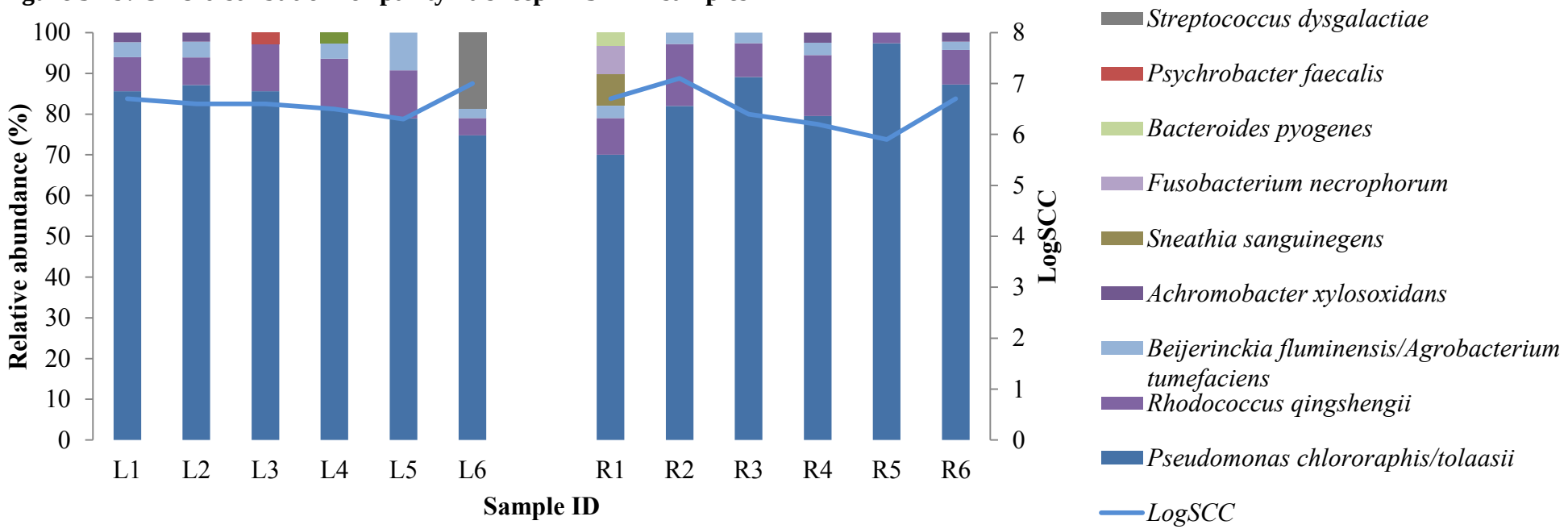


Figure 5-28: DGGE image of parity 4 sheep A40

Figure 5-29: OTU distribution for parity 10 sheep A23 milk samples



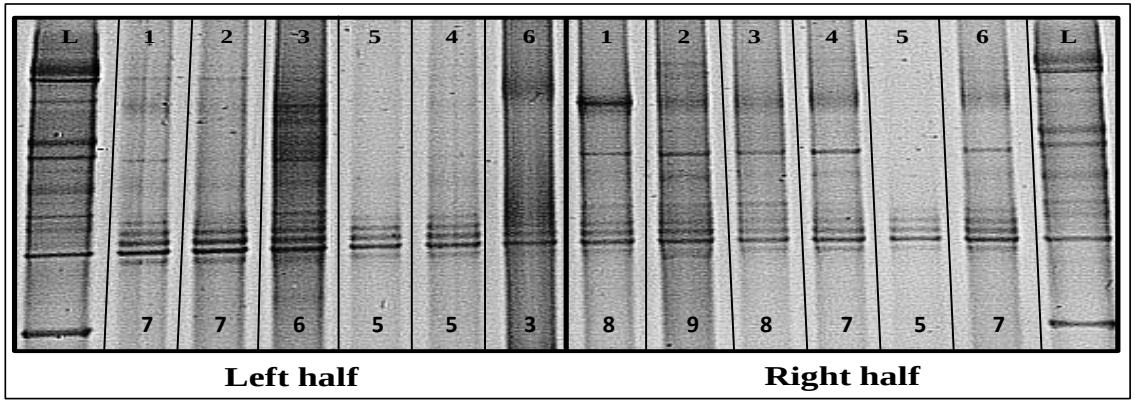


Figure 5-30: DGGE image for parity 10 sheep A23

Figure 5-31: Technical replicate comparison for sample L8 from parity 1 sheep A20

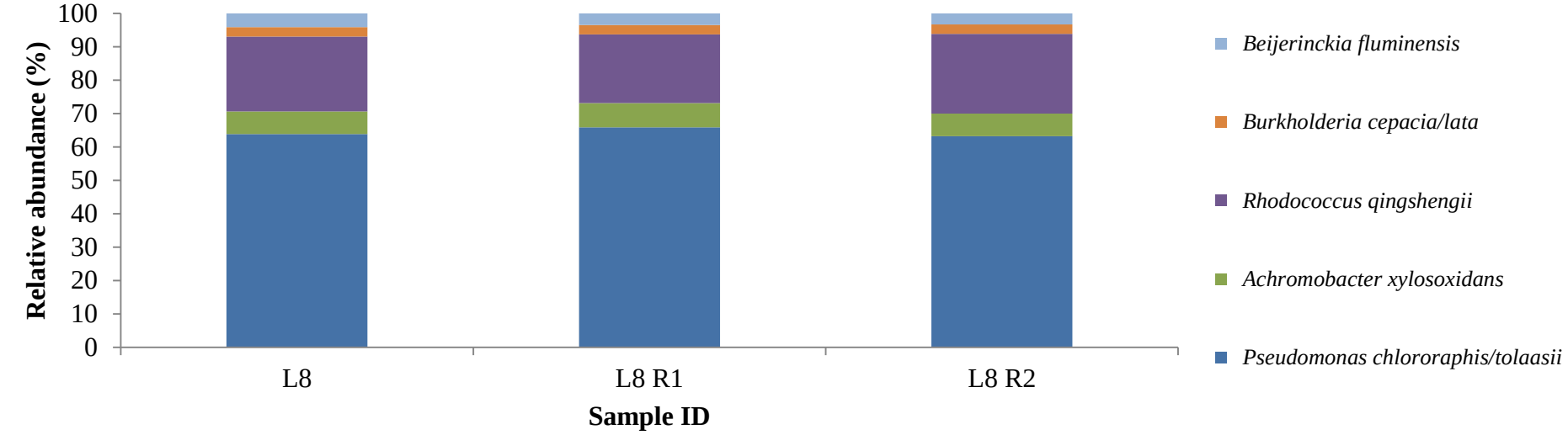
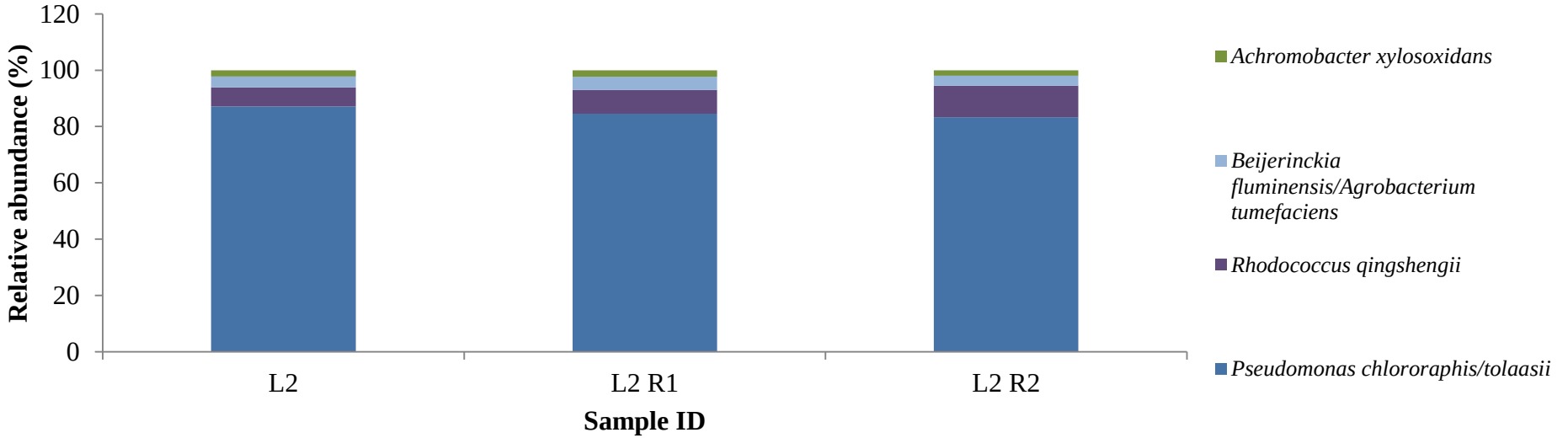


Figure 5-32: Technical replicate comparison for sample L2 of parity 10 sheep A23



5.4 Discussion

5.4.1 Library preparation

The protocol to prepare milk samples for MiSeq sequencing analysis required optimisation. This was necessary to ensure that the results were an accurate representation of the composition of each milk sample processed. As the starting template was PCR product as opposed to DNA, the number of PCR cycles for each step in the library preparation was extensively tested to achieve the balance between sufficient PCR product yield and minimal cycles to reduce formation of PCR artefacts and/or chimeras. Each step included a negative control and negative controls from each step in library preparation were included in subsequent steps to ensure any contamination could be detected. The final protocol produced is easily adaptable to other sample types for future projects and produces good quality data for analysis.

5.4.2 Data analysis pipeline

The methodology used to interpret sequencing data is critical as it can directly affect the conclusions drawn from a data set. Hence, each step and associated parameters chosen in the data analysis pipeline for this data set were carefully considered, with parameters used only when there was rationale to do so. For example, in the first step to merge the forward and reverse read for each sample, different quality truncation and mismatch scores were tested in a range of combinations for both a positive control and a milk sample to assess the effect of both parameters before determining which to use. To ensure only good quality data were analysed, stringent parameters were chosen where possible, resulting in only 1 mismatch and a truncation score of 3, in comparison to the default of any number of mismatches and no truncation (Section 5.3.3, Table 5-9 and Table 5-10).

Data were quality filtered based on the maximum error rate and sequence length. The maximum error rate was selected instead of the average quality score as it determines the probability of each base in a read being incorrect to give the most likely number of errors across a read. This is a more accurate interpretation of read quality in comparison to taking an average. Furthermore, rather than arbitrarily selecting a sequence length cut-off, a ReLIC analysis was conducted (Williams and Purdy, submitted). A ReLIC analysis is based on the hypothesis that at a certain point along the read length distribution, a transition between true biological diversity and polymorphisms arising from sequencing errors occurs. In the

sequencing analysis, individual sample files were clustered into OTUs using CD-HIT (Li and Godzik, 2006). The ratio of the increase in the number of OTUs over the increase in the number of sequences within each sample was calculated to provide the cumulative increase in the number of reads and OTUs incrementally away from the median read length. A marked increase in the ReLIC value below 312bp indicated that reads below this length did not cluster with either existing clusters or form new abundant clusters at that length, suggesting reads below this length were of poor quality and hence were excluded from further analysis.

Each library contained a control sample of a model community consisting of 5 known bacterial species commonly associated with intramammary infections. The use of a model community provides another method for filtering sequencing data to remove sequences that are potentially chimeras or PCR artefacts that may distort the assessment of sample diversity and composition. The clustering of a model community from each library resulted in the 5 expected OTUs, with 9 additional OTUs (Table 5-15), 8 of which had a BLAST identification of *Escherichia coli*, although the sequence coverage and identity confidence matched other, unrelated bacterial species. These results suggested that erroneous sequencing reads were forming additional OTUs and although these could be differentiated in the model community, this would not be the case for study samples.

Hence, in order to address the study hypotheses with confidence in the supporting sequencing data, the relative size of the largest erroneous OTU in each dereplicated model community was determined and the largest overall erroneous OTU used as a size marker to filter the sequencing data. This does introduce the potential for rare OTUs that are real components of the community to be discarded, although these reads cannot be easily distinguished from sequencing errors and chimeras and so cannot be analysed with confidence. More conservative but more realistic conclusions therefore can be drawn from data that can be considered reliable based on the OTU filtering process.

5.4.3. Interpretation of MiSeq results of 5 sheep

5.4.3.1 Parity 1 sheep A20

Sheep A20 had 14 milk samples; 7 per mammary gland half, with 10 OTUs (Figure 5-21). The DGGE for sheep A20 (Figure 5-22) showed a good overall level of correlation with the number of OTUs. For example, week 2 of the left half had both 8 DGGE bands and 8 OTUs and the DGGE showed that week 2 was different from weeks 3-8, which is supported by the OTU distribution in Figure 5-22. Also, for weeks 3-8 of the left half, the DGGE banding

pattern was identical, with 6 DGGE bands detected in each sample. This is corroborated by the OTU distribution which showed the same 5 OTUs in samples from weeks 3-8. Even though the DGGE indicated 6 bands for the 5 OTUs, the consistent DGGE banding pattern does show that these samples have a similar composition in agreement with the MiSeq data.

For the right half of sheep A20, weeks 2, 3 and 6 have the same number of OTUs and DGGE bands. However, there are some differences. The DGGE indicated week 2 as having a different composition, with a DGGE band not seen in any other sheep A20 sample. The OTU distribution showed that *Corynebacterium efficiens* is present only in weeks 1 and 2, possibly indicating that the DGGE detected this in week 2 but not in week 1. The DGGE pattern for week 1 of the right half has some smearing, which may have made band identification more difficult.

The DGGE suggests weeks 1, 3, 4, 5 and 6 of the right half of A20 are similar. Weeks 4 and 5 are similar in terms of bacterial species detected, but weeks 3 and 6 have *Pseudomonas olevorans* not found in weeks 4 and 5 and week 1 contains *Corynebacterium efficiens* and *Psychrobacter faecalis* which are not seen in weeks 3-7. Week 7 had a similar OTU distribution but different relative abundance to weeks 4 and 5, but 1 more DGGE band. Hence, the overall patterns indicated in the DGGE are supported by the sequencing results, although the sequencing provides a more in-depth identification of members of the bacterial community.

Of the 10 OTUs detected in sheep A20, 7 had an association with change in SCC in Section 4.3.4 of Chapter 4. Both *Rhodococcus qingshengii* and *Achromobacter xylosoxidans* were previously associated with a lower SCC and *Psychrobacter maritimus*, *Burkholderia cepacia*, *Corynebacterium efficiens*, *Psychrobacter faecalis* and *Pseudomonas chlororaphis* associated with a significantly higher SCC. Interestingly, the SCC does decrease between weeks 1 and 2 of the left half, with 3 bacterial species associated with a higher SCC present in week 1 and not in week 2. However, the SCC also decreased between weeks 1 and 2 of the right half, despite the presence of 2 bacterial species associated with a lower SCC in week 1. This could suggest that differences in community composition other than the presence of the bacterial species associated with a change in SCC in the model may be having an effect on SCC.

Overall, a persistent community was detected in milk samples from sheep A20. Earlier weeks (i.e. week 1 and 2) in lactation had the most diverse composition, with the remaining weeks showing similar stable community compositions, with fluctuations in relative abundance

from week to week. Both mammary gland halves had similar compositions, with some differences e.g. *Pseudomonas olevorans* was only present in the right half. Furthermore, differences were also related to specific time points. For example, *Psychrobacter maritimus* was only found in week 1 of the left half. *Pseudomonas chlororaphis* dominated in all sheep A20 milk samples. This is in agreement with a study by Kuehn *et al.*, (2013) which found *Pseudomonas spp*. in both healthy and diseased milk samples from dairy cattle, but with dominance in healthy milk samples.

5.4.3.2 Parity 2 sheep A41

For sheep A41, 11 milk samples were processed; 6 from the left half and 5 from the right, producing 34 OTUs (Figure 5-23), many more OTUs than the 4 other sheep presented. However, Kuehn *et al.*, (2013) found an average of 30 OTUs per healthy milk sample, although the data pipeline to analyse this data was different. Despite this, such data does indicate the potential diversity of the mammary gland microbiome.

Comparing the DGGE profile for sheep A41 (Figure 5-24) to the OTU distribution, the DGGE does show a more complex and diverse community in comparison to the other 4 sheep, which is in agreement with the larger number of OTUs per sheep A41 milk sample. However, the DGGE does identify more bands then there are OTUs. For example, in the left half, from weeks 1-6, 8, 10, 8, 10, 17 and 9 OTUs were identified, yet the same samples in the DGGE have 16, 21, 18, 21, 19 and 21 DGGE bands respectively. It may be that uncharacterised artefacts and/or chimeras from the PCR process (Delgado *et al.*, 2013; Giraffa and Neviani, 2001) have distorted the view of community diversity. Also, 1 DGGE band does not always correlate to 1 bacterial species, as previously discussed in Section 4.4.3 of Chapter 4. In addition, the DGGE banding pattern in Figure 5-24 is compact, which may have resulted in a single band being identified as more than one in the Gel Compar II analysis.

Despite this, the DGGE does indicate that sheep A41 had a more diverse bacterial community in comparison to the other 4 sheep analysed with some correlations visible. For example, week 3 of the left half had a distinct band (highlighted in Figure 5-24) and this sample is dominated by *Mannheimia haemolytica*. A band of lower intensity in the same position is visible in week 4 and 5 samples, which also contain *Mannheimia haemolytica*.

Only 6 of the 34 OTUs were significantly associated with a change in SCC (Section 4.3.4 of Chapter 4) *Mannheimia haemolytica*, *Atopostipes suicloacalis* and *Jeotgalicoccus*

psychrophilus were associated with a lower SCC and *Psychrobacter maritimus*, *Corynebacterium efficiens* and *Psychrobacter faecalis* with a higher SCC. *Mannheimia haemolytica* was present in weeks 2, 3 and 4 of the left half, with SCC rising over these weeks, although *Mannheimia haemolytica* was also present in a lower abundance in week 5 where there was a reduction in SCC. This suggests that bacterial pathogens can be present when there are no clinical signs of disease as part of a commensal microbial community.

Atopostipes suicloacalis was present in all but week 3 of the left half and week 6 of both halves. Overall, when *Atopostipes suicloacalis* was not present the SCC increased, apart from between weeks 2 and 3 of the right half, which may suggest a potential protective role. However, the rise in SCC in the absence of *Atopostipes suicloacalis* does not confirm cause and effect, as the rise in SCC could be due to other bacterial species and/or changes in the community unrelated to *Atopostipes suicloacalis*.

Jeotgalicoccus psychrophilus was only present in week 6 of the right half; its effect on SCC is difficult to decipher as this sample is the most diverse, with 15 OTUs. *Psychrobacter maritimus* was only present in week 2 of the right half, which had a median SCC in comparison to the SCCs for other sheep A41 samples. Despite being associated with a higher SCC, *Psychrobacter faecalis* was present in all samples apart from week 6 samples where the SCC was highest. Kuehn *et al.,* (2013) found *Psychrobacter* spp. only in milk samples from healthy dairy cattle. These findings suggest *Psychrobacter* spp. are commensal organisms in healthy mammary glands.

Corynebacterium efficiens dominates milk samples from sheep A41, with its detection in all of the milk samples, with varying relative abundance over lactation. Interestingly, its lowest abundance is in week 3 of the left half, where *Mannheimia haemolytica* dominates. This could suggest there is competition between bacterial species within the mammary glands. *Corynebacterium* spp. found in 4 of the 5 sheep in this study have been found in both healthy and mastitic milk samples from dairy cattle (Kuehn *et al.*, 2013), suggesting the genera can be present regardless of disease state.

Furthermore, *Staphylococcus equorum* was detected in 80% of sheep A41 milk samples, and a study by Oikonomou *et al.,* (2012) found *Staphyloccocus* spp. in mastitic and healthy milk. A study by Oikonomou *et al.,* (2014) identified *Staphylococcus aureus* in small quantities in healthy quarters of dairy cows. Hunt *et al.,* (2011) found a dominance of *Staphyloccocus* spp. in non-clinical human breast milk. In this study, *Staphylococcus* spp.

were only detected in sheep A41. Despite the findings in cow and human milk, the low abundance of *Staphyloccocus* spp. in the milk samples from 4 of the 5 sheep investigated could suggest they might not be persistent in the commensal bacterial community of healthy sheep, although analysis of data from the remaining sheep in the study is required to test this hypothesis.

Sheep A41 shows a persistent microbial community with changes in composition and abundance of bacterial species over lactation. Despite the complexity of the community in sheep A41, there were similarities in community composition over lactation and half. For example, *Psychrobacter sanguinis*, *Corynebacterium efficiens* and *Facklamina tabacinasalis* were present in all samples. However, there were also differences within halves e.g. *Staphylococcus aureus* was only present in week 4 of the right half and *Acidovorax ebreus* only in week 5 of the left half. Furthermore, it could be hypothesized from sheep A41 that parity 2 sheep have a more complex bacterial community in comparison to parity 1. This provides evidence to support the hypothesis that the complexity of the community increases with age, although analysis of more sheep of differing parities is required to confirm this.

5.4.3.3 Parity 3 sheep A25

Parity 3 sheep A25 had 14 milk samples processed; 8 from the left half and 6 from the right half, with 10 OTUs identified (Figure 5-25). Of these 10 OTUs, 7 were shared with parity 1 sheep A20, indicating similarities in bacterial community composition between sheep. Like sheep A20, sheep A25 samples were dominated by *Pseudomonas chlororaphis* which was associated with a higher SCC in the model.

Interestingly, *Burkholderia spp.*, *Brevundimonas spp.*, and *Escherichia coli* have been associated with both subclinical and clinical mastitis in previous studies in dairy cattle (Bhatt *et al.*, 2012; Kuehn *et al.*, 2013; Oikonomou *et al.*, 2014). Despite this, they are all present in sheep A25 which is non-diseased. This again suggests that bacterial pathogens can be present when there is no clinical disease. It could therefore be hypothesized that these opportunistic pathogens cause disease when there are changes in community diversity as opposed to presence alone resulting in disease.

The DGGE of sheep A25 (Figure 5-26) does show a correlation to the OTU distributions. For example, weeks 3, 5, 6, 7 and 8 of the left half all contain the same 6 OTUs and the DGGE banding pattern for these samples is identical, although only 3 DGGE bands were identified per sample.

Furthermore, the DGGE shows a change in banding pattern between the left and right halves which is visible in the OTU distribution. All but 2 of the left half samples contained *Achromobacter xylosoxidans* which was only found in weeks 6 and 7 of the right half. *Achromobacter xylosoxidans* has been associated in modelling work with a lower SCC and the milk samples in the left half of sheep A25 do have lower SCCs in comparison to the right half. The right half contains *Escherichia coli*, *Corynebacterium efficiens*, *Psychrobacter faecalis* and *Elizabethkingia miricola*, none of which are found in the left half. Both *Corynebacterium efficiens* and *Psychrobacter faecalis* were significantly associated with a higher SCC in the model, which agrees with the higher SCCs seen in the right half. This indicates that the sequencing data could provide evidence to support the associations with SCC found in the model in Chapter 4.

There were also similarities across sheep A25 milk samples. *Pseudomonas chlororaphis*, *Rhodococcus qingshengii*, *Burkholderia* spp. and *Beijerinckia fluminensis* / *Agrobacterium tumefaciens* were found in all samples. It is interesting that these bacterial species were present in both halves, yet the SCC in the right half was higher. This could suggest that the combination of bacterial species and/or interactions that occur in the right half are linked to a higher SCC.

Overall, sheep A25 shows a persistent bacterial community with similarities in composition over lactation and between halves. However, there were also differences between the bacterial community and SCC between mammary gland halves. Sheep A25 was more similar in community composition to parity 1 sheep A20 than parity 2 sheep A40. Despite the increase in the number of bacterial species with increasing parity from parity 1 to parity 2, an equivalent increase from parity 2 to 3 was not seen. Therefore, the data presented does not support the hypothesis that the number and species of bacteria colonising the mammary gland will increase over lactation and with sheep age. However, as the data presented contains only 1 sheep per parity, further analysis is required to elucidate the accuracy of this hypothesis.

5.4.3.4 Parity 4 sheep A40

There were 13 milk samples processed for parity 4 sheep A40; 6 from the left mammary gland and 7 from the right, with 6 OTUs identified (Figure 5-27). All of these OTUs were identified in parity 1 sheep A20 and parity 3 sheep A25, with 2 (*Corynebacterium efficiens* and *Psychrobacter faecalis*) also identified in parity 2 sheep A41. *Pseudomonas chlororaphis* was the most abundant bacterial species in all sheep A40 milk samples, as was the case for

sheep A20 and A25. Of the 6 OTUs in sheep A40, 5 were identified in the model, with 3 significantly linked to a higher SCC (*Pseudomonas chlororaphis*, *Corynebacterium efficiens* and *Psychrobacter faecalis*) and 2 associated with a lower SCC (*Achromobacter xylosoxidans* and *Rhodococcus qingshengii*).

The bacterial species associated with a lower SCC were present in all milk samples apart from week 2 of the right half (*Achromobacter xylosoxidans* not present). Interestingly, both *Corynebacterium efficiens* and *Psychrobacter faecalis* were only present in week 4 of the right half, which has the second lowest SCC, despite both of these species being associated with a higher SCC in the model. It could be possible that in this milk sample, other dominant commensal bacterial species present in large abundances such as *Pseudomonas chlororaphis* have had a protective effect, suppressing SCC.

Similarly, the difference in community composition in week 4 of the right half was detected in the DGGE for sheep A40 (Figure 5-28) which contained 2 additional DGGE bands not seen in any of the other milk samples. The DGGE also indicated a similar community across the rest of the milk samples, which was also seen in the OTU distribution, with 4 OTUs present in every sample apart from week 2 of the right half (3 OTUs).

Overall, sheep A40 had a similar bacterial community across both mammary gland halves, with some fluctuations in the relative abundance of the 4 dominant OTUs, indicating a stable community over time which correlates with results seen for human breast milk (Hunt *et al.*, 2011). Despite this, differences in bacterial community were present within the right mammary gland half. Parity 4 sheep A40 had a similar bacterial community to parity 3 sheep A25 and parity 1 sheep A20. It could be hypothesized that parity 3 and 4 sheep have more stable communities after exposure to a range of bacterial species in earlier years, with parity 2 sheep A41 illustrating the transition between a developing and a stable microbiome. Such exposure may require time, suggesting why parity 1 sheep do not have the same community complexity as parity 2.

5.4.3.5 Parity 10 sheep A23

There were 12 milk samples processed for parity 10 sheep A23; 6 per mammary gland half, with 9 OTUs identified (Figure 5-29). Of these 9 OTUs, 5 were in common with parity 3 sheep A25 and parity 4 sheep A40, 6 were in common with parity 1 sheep A20 and 1 (*Psychrobacter faecalis*) with parity 2 sheep A41. As seen in all sheep apart from A41,

Pseudomonas chlororaphis was most abundant in all of sheep A23 milk samples, indicating similarities in community composition between halves and sheep.

The dominance of *Pseudomonas chlororaphis, Rhodococcus qingshengii* and *Beijerinckia fluminensis*/ *Agrobacterium tumefaciens* in milk samples across sheep in the study with no clinical signs of disease could suggest these are actually commensal organisms that have a role in maintaining stability in the mammary gland microbiome. A study of human breast milk by Hunt *et al.,* (2011) identified a 'core' microbiome of 9 OTUs present in 16 women. In this study, 3 OTUs (*Corynebacterium efficiens, Psychrobacter faecalis* and *Beijerinckia fluminensis*) were found in all milk samples from 4 of the 5 sheep investigated and like the study by Hunt *et al.,* (2011), these 3 OTUs were found to represent a large proportion of the abundance of the bacterial communities observed.

The DGGE image for sheep A23 (Figure 5-30) showed a difference in banding pattern in week 1 of the right half which correlates with the OTU distribution, with this milk sample having the largest number of OTUs and therefore the most diverse bacterial community. Weeks 2, 3 and 4 of the right half have similar DGGE banding patterns and OTU distributions, although week 4 does have *Achromobacter xylosoxidans* which is not seen in weeks 2 and 3. The number of DGGE bands in week 5 of the right half was the lowest in the half and this corresponds to the OTU profile, which is also the simplest, with 2 OTUs. For the left half, weeks 1, 2 and 4 all have the same 4 OTUs and a similar DGGE pattern, although 2 less DGGE bands are recorded in week 4. The intensity of the DGGE banding pattern changes in weeks 3 and 6 of the left half, which corresponded to changes in the bacterial species present in these samples in comparison to the week before.

Differences between the bacterial communities according to mammary gland half and time were also visible in OTU distribution. For example, week 3 of the left half contained *Psychrobacter faecalis*, which is not seen in any other sheep A23 milk sample, as is the case for *Streptococcus dysgalactiae* in week 6. In the right half, week 1 contained *Fusobacterium necrophorum, Bacteroides pyogenes* and *Sneathia sanguinegens*, also not seen in any other milk samples.

Studies by Oikonomou *et al.,* (2014; 2012) reported a high abundance of anaerobic bacteria such as *Fusobacterium* spp. in both clinical and healthy samples. In this study, only 1 sample from sheep A23 contained *Fusobacterium necrophorum* and the anaerobe *Bacteroides pyogenes* was detected in 2 milk samples; 1 in sheep A20 and 1 in sheep A23. Oikonomou *et*

al., (2012) suggests that the role of anaerobes has been underestimated due to the use of bacterial culture as the gold standard to identify causative agents of mastitis.

Interestingly, *Fusobacterium necrophorum* is detected in the only milk sample across the 5 sheep that also contains *Sneathia sanguinegens*. This could suggest that anaerobes like *Fusobacterium* spp. may form synergistic relationships with other bacterial species such as *Sneathia sanguinegens* and so are more likely to be present in association with one another. This idea is plausible as a synergistic relationship between *Fusobacterium necrophorum* and *Trueperella pyogenes* in the aetiology of summer mastitis has previously been reported (Panciera *et al.,* 1989). Furthermore, Oikonomou *et al.,* (2012) found *Sneathia* spp. but only in samples that were classified as either *Escherichia coli*, *Trueperella pyogenes* or *Klebsiella* spp. in classical bacteriology. This could indicate that *Sneathia sanguinegens* forms mutually beneficial interactions with pathogens, assisting them in becoming more dominant community members.

Overall, parity 10 sheep A23 shows a persistent bacterial community consisting of several OTUs shared by other sheep. Similarities and differences over time and according to mammary gland half were detected and potentially synergistic relationships between certain bacteria species were hypothesized.

5.4.4 Summary of MiSeq results of 5 sheep

The MiSeq sequencing data presented shows 49 OTUs from 65 sheep milk samples (33 from the left and 31 from the right half) for 5 sheep. The data presented in this Chapter only consists of 1 sheep per parity group and only 5 of the 30 sheep in the study. Hence, the discussion based on these data is highly speculative, only providing inferences on the validity of the study hypotheses.

In 4 of the 5 sheep, 3 OTUs were found in all milk samples. Oikonomou *et al.,* (2014) found 4 bacterial genera in every sample obtained from a healthy quarter in dairy cattle. In this study, between 2 and 15 OTUs were identified per milk sample, suggesting a blend of bacterial species are present in the mammary gland, regardless of SCC or disease state. Both Oikonomou *et al.,* (2014) and Bhatt *et al.,* (2012) reported a mixture of bacterial species in clinical and subclinical milk from dairy cattle. Hence, the results of this study suggest colonisation of the mammary gland is inevitable with a microbial community present and persisting in every milk sample.

The results presented in this Chapter validate the use of DGGE to visualise community diversity. Several of the bacterial species associated with a change in SCC identified in Chapter 4 have also been detected in the data from the 5 sheep presented. At times, the number of DGGE bands and OTUs were an exact match e.g. left half; week 1; parity 1 sheep A20 or within 1 e.g. left half samples for sheep A40. Despite some underestimation of diversity at times e.g. left half; week 1; sheep A25 and overestimation e.g. parity 2 sheep A41 samples, changes in DGGE patterns corresponded well to changes in OTU distribution. Often, one or more bacterial species associated with a change in SCC in Chapter 4 were present when there was a change in community composition in Chapter 5. Therefore, the conclusions based on DGGE data in Chapter 4 remain valid. Despite this, the sequencing data undoubtedly offers a more in-depth and comprehensive view of the ecology of sheep milk bacterial communities and so further analysis of data from all 30 sheep is required to effectively address the study hypotheses.

Whilst the results presented are only from 5 sheep, the hypothesis that the number and species of bacteria would increase with increasing parity is not illustrated in the data presented. Bacterial communities from parity 1, 3, 4 and 10 sheep were relatively stable with 6-10 OTUs identified and 3 OTUs (corresponding to *Pseudomonas chlororaphis*, *Rhodococcus qingshengii* and *Beijerinckia fluminensis* / *Agrobacterium tumefaciens*) present in all milk samples. In contrast, parity 2 sheep A41 showed a more complex community composition with 34 OTUs, with only 3 of these OTUs found in milk samples from the other sheep. This could be representative of the difference in bacterial community across sheep according to parity or indicative of a personalised community specific to sheep A41 as community profiles specific to an individual animal have been previously reported (Kuehn *et al.*, 2013).

Parity 3 and 4 sheep were relatively stable over time, with the same OTUs in multiple milk samples across both mammary gland halves which has previously been reported in human milk (Hunt *et al.*, 2011). This could represent a stable mammary gland microbiome, as the sheep are of an age where they are likely to have been exposed to a wide range of bacteria. Parity 2 sheep could be in a transitional phase where bacterial species are competing for domination. Parity 1 sheep may have a less complex community due to the time it would take to accrue the complex community seen in parity 2. Parity 10 sheep A23 had a greater number of OTUs than parity 4 sheep A40 and the same number of OTUs as parity 3 sheep A25. As mammary gland defences may deteriorate with age as previously discussed (Chapter 4

Section 4.4.3), it could be expected that older sheep may see an increase in community complexity. Alternatively, changes in community composition over time may be random and so correlations according to sheep age may be difficult to elucidate.

Differences in bacterial community according to mammary gland half were detected, most clearly in parity 3 sheep A25, where 4 OTUs were specific to right half samples. More commonly in this data set, differences in mammary gland half were due to changes in bacterial community composition at specific time points. For example, week 4 of the right half for sheep A40 had 2 OTUs not seen in any other milk samples for sheep A40. Changes over time in different quarters of the same dairy cattle have been previously reported (Kuehn *et al.*, 2013). This suggests that there can be changes in community composition over lactation which are specific to a mammary gland half. Such changes could be random or linked to factors specific to a mammary gland half such as udder conformation. As the two halves are separate entities, differences in community composition would be expected.

However, similarities in community composition over time, mammary gland half and sheep were also detected. This could suggest that bacterial species most adapted to the mammary gland environment persist, and so similarities in community will be detected across sheep on the same farm.

Bacterial species previously associated with both healthy and clinical mastitis milk samples have been identified. For example, *Psychrobacter* spp. were found in all 5 sheep and this genera has been found in healthy milk samples from dairy cattle (Kuehn *et al.*, 2013). *Staphylococcus* spp. were found in sheep A41 and this genera has been identified in both healthy milk in humans (Hunt *et al.*, 2011) and both healthy and diseased milk samples in dairy cows (Oikonomou *et al.*, 2014). Furthermore, some bacterial genera associated with clinical disease in dairy cattle such as *Sphingomonas* spp. and *Stenotrophomonas* spp. (Kuehn *et al.*, 2013) have not been found in the 5 sheep investigated. Oikonomou *et al.*, (2012), reported a large abundance of anaerobic bacteria in mastitic milk from dairy cattle such as *Trueperella pyogenes*, a bacterial species not detected in the sheep analysed so far in this study. Therefore, the observations in this study could add support to previous findings that healthy and diseased milk samples have differing microbiota profiles (Kuehn *et al.*, 2013; Oikonomou *et al.*, 2014). It must be noted however, that differences between this and other milk microbiome studies could be related to a combination of differences in sample

processing, farm environment and management and/or sequencing method/data analysis pipeline.

Furthermore, bacterial pathogens associated with causing intramammary infection were detected in milk samples from healthy sheep. This could suggest that a commensal community consisting of both pathogenic and non-pathogenic bacterial species is possible. Changes as a result of factors related to individual sheep, the surrounding environment or management could then predispose an animal to infection. As in some cases there was a change in SCC but a consistency in community composition, the interactions between pre-existing bacterial species within a community may play an important role in determining whether infection occurs as opposed to the introduction of a new species resulting in infection.

The sequencing results in this Chapter indicate a persistent bacterial community in non-diseased sheep, with changes over time, mammary gland half and sheep age with no OTU present in all 5 sheep. Hence, these results suggest the hypotheses of a natural microbial community forming in the sheep mammary gland and sheep-specific communities forming may be correct. As changes in community composition and SCC are seen without the development of clinical signs of disease, the results for these 5 sheep suggest that changes in community diversity alone may not be sufficient to trigger infection. As approximately 20% of milk samples in the study overall have SCCs that would be considered clinical in dairy cattle, it may be that changes in community seen in some study samples could be resulting in subclinical disease. However, further analysis of data from all 30 sheep is required to determine the accuracy of the study hypotheses.

5.4.5 Future work and conclusions

Due to time constraints, only data from 5 sheep has been presented. Further data processing and analysis is required to investigate the study hypotheses and new hypotheses generated from using data from all 30 sheep. To do this, the filtering process must be completed. Currently, each sample in the study has a mapping file that details the sequences from each sample that map to filtered OTUs and those that do not. Hence, custom Perl scripts need to be developed to use the mapping file to remove sequences that do not map to filtered OTUs from the quality filtered file for each sample. With unreliable sequences removed, all of the study data can then be dereplicated and clustered to facilitate a comparison across all 30

sheep. The clustered dataset could then be explored further to address the study hypotheses regarding changes in community composition with parity, lactation and mammary gland half.

OTUs from the global clustering process could also then be incorporated into the mixed effects regression model to investigate the effect of specific bacterial species (OTUs) and different combinations of bacterial species on SCC. This process would not only enable community composition to be investigated further, but also the interactions between different bacterial species, which is essential in understanding how intramammary infections develop.

Previous studies have indicated differences in community composition according to disease state (Kuehn *et al.*, 2013; Oikonomou *et al.*, 2014). To elucidate further the changes in bacterial community when clinical infection develops, future longitudinal studies would benefit from sampling the transition from a healthy to clinically diseased state. Furthermore, as community composition is likely to have an environmental influence, investigation of sheep on more than one farm would enable OTUs that are specific to individual farms to be investigated. This in turn could enable the development of farm-specific management measure to reduce rates of intramammary infection.

In conclusion, the data presented in this Chapter indicates that a persistent and diverse community is present in the healthy sheep mammary gland. Weekly fluctuations in relative abundance of bacterial species are indicated, with both similarities and differences between mammary gland halves and time points. Similarities in bacterial community composition between different sheep of differing parities have been identified, as well as sheep specific community compositions. Bacterial species associated with both healthy and clinical milk samples have been found, indicating that bacterial pathogens could be present as part of a natural microbial community when there are no clinical signs of disease. A comprehensive analysis of the data from all study samples will help to elucidate the study hypotheses further.

Chapter 6 : General discussion and future directions

6.1 Introduction

The overall aim of this study was to obtain an understanding of the bacterial genera present in the microbial community of the suckler sheep mammary gland using molecular-based whole community approaches that were culture-independent.

In order to address this aim, a longitudinal study was designed and the following hypotheses were established:

1. A natural microbial community (microbiome) forms in the suckler sheep mammary gland.

2. Perturbations in the community result in infection and disease (change in SCC).

3. With increasing number of lactations, the number and species of bacteria colonising the mammary gland increases.

4. Differences in microbial community composition occur between mammary gland halves.

5. Colonisation of the mammary gland is inevitable.

Evidence to support and/or speculate on the validity of these hypotheses has been provided by this study. A summary of this evidence with a discussion of its limitations and further work required are discussed in this Chapter.

6.2 Research findings and implications

A culture-independent approach was chosen to assess the microbial community in suckler sheep mammary glands. The first step in using this approach was DNA extraction. The results in Chapter 2 highlight the importance of thoroughly testing different DNA extraction methods in order to select the most appropriate method for the sample type and subsequent downstream analysis. Chapter 2 also identified intermittent contamination in kit-based DNA extraction methods. This indicates the importance of using negative controls to determine the DNA extraction method that produces the most representative results. This is an important consideration in microbiome studies, where even trace amounts of contaminants may produce significant numbers of reads in high-throughput sequencing analyses. The selected DNA extraction method in Chapter 2 consistently amplified DNA from low bacterial yield milk samples without the detection of contamination. Hence, this method has great potential for use in further microbiome studies, with research already in progress to enhance the high-throughput capability of the method for larger longitudinal studies.

The second step in sample processing was PCR-DGGE. In Chapter 3, 18 PCR primer sets, 3 PCR master mixes and the DGGE protocol were tested to identify the optimal PCR-DGGE approach for this study. The results from Chapter 3 show how different PCR primers, master mix combinations and PCR conditions can affect PCR product yield, highlighting the importance of testing and optimising different PCR protocols to identify the most effective combination. Chapter 3 also shows the importance of optimising the PCR protocol with the GC clamp attached to ensure that sufficient PCR amplification is achieved with the GC clamp for visualisation of the community using DGGE. The production of a custom DGGE reference ladder detailed in Chapter 3 would be recommended for future studies to ensure the accuracy of DGGE band identification, comparison and excision.

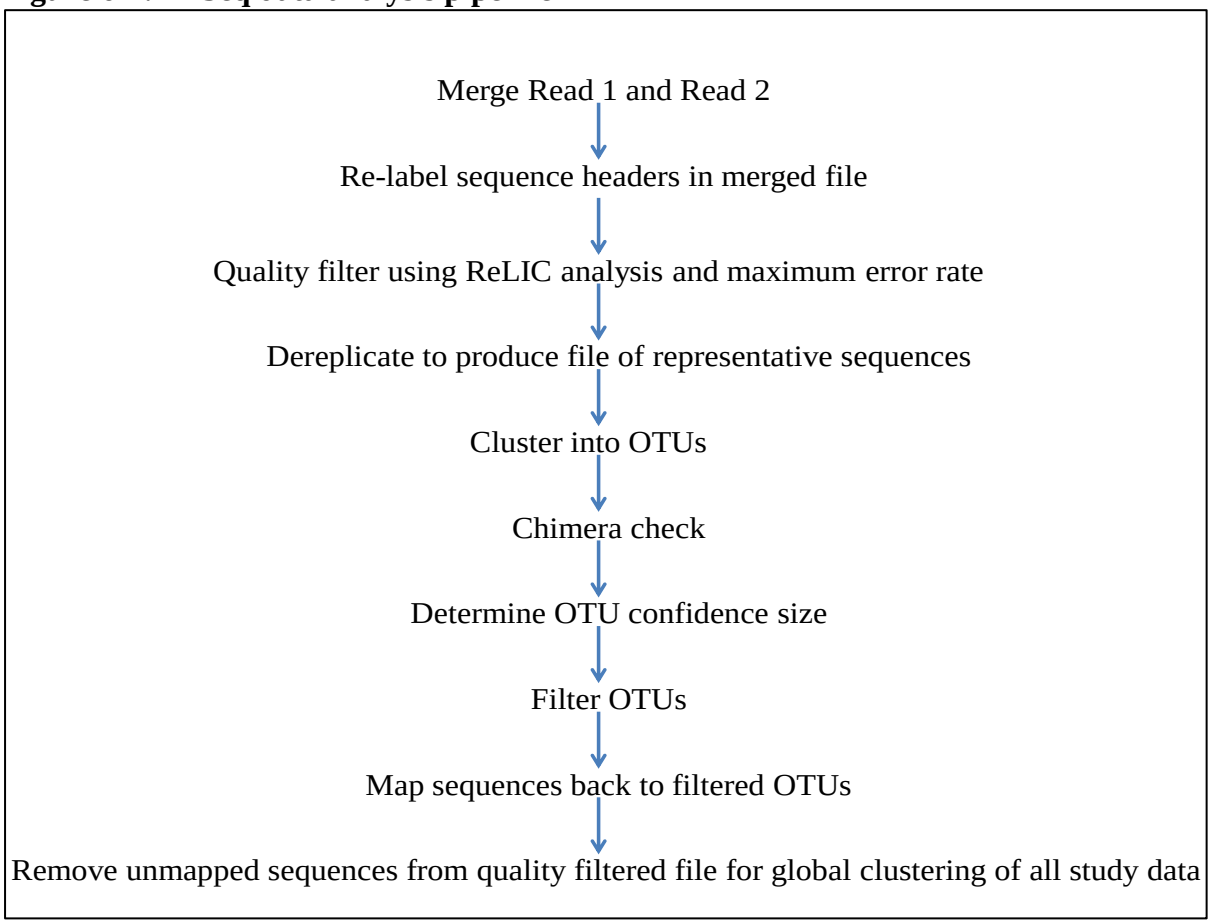
In Chapter 4, the methods optimised in Chapters 2 and 3 were used to initiate the investigation of the study hypotheses. The DGGEs produced in Chapter 4 identified 2-23 DGGE bands per milk sample. Even though a single DGGE band may not always correspond to a unique bacterial species, these DGGEs do provide evidence for the hypotheses that a microbial community is present in the sheep mammary gland and that there are both similarities and differences over lactation and between mammary gland halves and sheep. The results from Chapter 4 also suggest that colonisation of the mammary gland is inevitable, although validation of this hypothesis would require further longitudinal studies to investigate

how and when the mammary gland microbiome develops. The mixed effects regression model, with identification of bands significantly associated with a change in SCC presented in Chapter 4, provides powerful evidence for the hypotheses regarding links between specific bacterial species causing or protecting against infection. This in turn provides some evidence that changes in the community can predispose an animal to infection.

In order to provide clarity on the associations between specific bacterial species and SCC identified in Chapter 4 and to gain a more in-depth consensus of community composition and change, methods to sequence and analyse data from all study samples using the Illumina MiSeq were developed in Chapter 5. The custom library preparation protocol detailed in Chapter 5 is easily adaptable to other sample types and produces ample amounts of data for analysis. The protocol can therefore be used in future studies that use Illumina MiSeq.

The data analysis pipeline detailed in Chapter 5 uses an iterative approach to select data for analysis. This approach results in the analysis including only those OTUs for which there is evidence they form a true representation of community composition. Each step completed in the pipeline for each individual sample to date is summarised in Figure 6-1.

Figure 6-1: MiSeq data analysis pipeline



The OTU confidence size for this dataset was determined using the model community present in each of the 5 libraries. As detailed in Section 5.3.3 of Chapter 5, for each library, the size of the largest false OTU was determined based on the known composition of the model community. The largest overall unreliable confidence interval size was then used to filter the OTUs in each individual sample. Once reads were mapped back to the filtered OTUs for each sample, unmapped reads that did not map to confident OTUs were identified. Future work involves the removal of unmapped sequences from the quality filtered file for each individual sample using custom Perl scripts. Each quality filtered file can then be dereplicated and concatenated to produce a file of representative sequences for the entire dataset which can then be clustered into OTUs. The OTUs can then be re-populated to explore the study hypotheses further.

For the example of sheep A20 in Chapter 5, the OTU filtering process reduced the number of OTUs per sample from as much as 2331 OTUs to 7 OTUs. The limitation of this filtering method is that rare but real OTUs may be removed in the filtering process if the OTU size is below the confidence limit. However, this method does result in the analysis of only those OTUs that can be considered, with confidence, to form part of the community; alongside the stringent parameters chosen in the rest of the analysis pipeline, conclusions drawn from this data are therefore less likely to be skewed by errors arising from the processes the study samples have gone through.

Chapter 5 also presents evidence to suggest that the data analysis pipeline produces an accurate representation of the community. Figure 5-19 and Figure 5-20 in Chapter 5 show a sharp decline in OTU size after the OTU size cut-off. This indicates that the OTU filtering process is not removing OTUs based on minute differences in OTU size, suggesting that the process is not removing data arbitrarily. Furthermore, DGGE banding patterns for the 5 sheep presented in Chapter 4 also show agreement with the sequencing data in Chapter 5. For example, 11 of the OTUs identified for the 5 sheep presented in Chapter 5 are bacterial species associated with a change in SCC from the DGGEs in Chapter 4. In several instances, the numbers of DGGE bands and OTUs in a sample were identical or in close agreement; changes in DGGE patterns in Chapter 4 were shown to result in changes in OTUs in Chapter 5. This suggests that the filtering process is not removing OTUs that have a significant effect on SCC. Furthermore, the correlations between the results for both Chapters indicate that the conclusions made from the DGGE analysis in Chapter 4 remain valid.

The results presented in Chapter 5 allow speculation on the accuracy of the study hypotheses. However, as these results only represent a small proportion of the sequencing data, the conclusions drawn are speculative. Further analysis and exploration of the sequencing data is required to elucidate the accuracy of the proposed hypotheses.

Despite this, some inferences on the accuracy of the hypotheses suggested can be made as well as suggestions on new hypotheses for further investigation. For example, the sequencing results in Chapter 5 support the hypotheses that a microbial community forms in the suckler sheep mammary gland, as a persistent community was detected in every sample from the 5 sheep presented. The data shown also indicate that there are differences in community composition according to mammary gland half, lactation and sheep age. However, similarities were also found, with OTUs shared by several sheep e.g. 3 OTUs were found in all milk samples from 4 of the 5 sheep presented, as well as similarities over time and between mammary gland halves of the same sheep. It could therefore be hypothesised that there would be similarities in mammary gland microbiota between halves and sheep, as bacterial species most adapted to the mammary gland environment will survive and persist in the mammary gland.

Furthermore, bacterial pathogens with a known association with intramammary infection such as *Staphylococcus aureus* and *Escherichia coli* were identified in healthy sheep milk. It could be hypothesised that pathogenic bacteria can form part of a commensal microflora in the mammary gland and changes in the community can then predispose these bacterial species to cause infection. Also, genera such as *Pseudomonas* spp. and *Psychrobacter* spp. dominated in the majority of the healthy milk samples analysed in Chapter 5. It could therefore be hypothesised that these genera form part of a commensal bacterial community in the sheep mammary glands. However, all of the study data would need to be analysed to provide evidence to support this hypothesis and further longitudinal studies of sheep on different farms would be required to account for any farm effect on mammary gland microbiota composition.

Similarly, as Chapter 4 identified associations between community composition and changes in SCC supported by the sequencing results in Chapter 5, it could be hypothesised that the interactions between community members are important in determining the predisposition to infection. Synergistic relationships between bacterial species have been suggested in

published literature (Green *et al.*, 2005; Panciera *et al.*, 1989; Witcomb *et al.*, 2014) and could also play a role in determining what organisms persist in the mammary gland.

Chapter 5 also suggests that the number and species of bacteria colonising the mammary gland does not increase with sheep age. However, as only 1 sheep per parity was analysed, this conclusion may not stand when further sheep of differing parities are incorporated into the analysis, although if the MiSeq data reflects the DGGE results, this would continue to be the case. It is interesting to note however, that the parity 1, 3, 4 and 10 sheep seemed to have relatively stable communities, with similar OTU numbers and abundances over time, whereas the parity 2 sheep had variable numbers of OTUs per sample with shifts in abundance over lactation and mammary gland half. It was proposed in Chapter 5 that parity 3, 4 and 10 sheep have more stable communities as their microbiomes have reached a state of equilibrium, with parity 2 representing the transitional period where bacteria are competing for dominance. However, as there is no clear pattern according to sheep parity from the data presented thus far, it is likely that community composition does not simply increase in complexity with age, as it is determined by several factors including those associated with individual animals, the bacteria, environment and management, making its development complex to elucidate and associate with age.

Furthermore, the hypotheses that the number and species of bacteria colonising the mammary gland will increase with time may be flawed. The risk of infection may increase over time as the immune system may deteriorate, or conformational changes in the udder make it easier for pathogens to invade at random and cause transient infections (Green *et al.*, 2005). However, this may not necessarily mean that the number of colonising bacterial species increases; changes in the community could result in commensal organisms causing infection or bacterial species may come and go at random as opposed to there being an association with time.

The identification of both similarities and differences over time, mammary gland half and sheep age could suggest that the composition of the mammary gland microbiome and what bacteria dominate could be determined at random, with different bacterial species appearing and disappearing over time by chance. Alternatively, further data analysis may suggest that the species most adapted to exploit the unique niche of the mammary gland are those that survive and persist to form a stable commensal community.

In relation to this, if future studies investigate sheep on different farms, it may be that the organisms that dominate are different between farms, as the species that may enter the mammary gland could differ according to factors such as environment and management. In turn, this could mean those that dominate do so by chance. If certain species consistently dominate, it could mean that they are more adapted to the mammary gland environment and so are more likely to persist once they gain access to the mammary gland.

6.3 Study conclusions

This is the first longitudinal microbiome study of intramammary infections in a farm animal. The data presented in this thesis provide the first evidence of a microbial community in the sheep mammary gland. In addition, the careful use of negative controls and the first description of a pipeline to rationalise the data can be used to elucidate the complex interactions between the sheep mammary gland microbiome and SCC. A persistent community has been detected over time, with similarities and differences by mammary gland half, lactation and age. Associations between certain community members and mammary gland health have also been identified through mixed effect modelling.

References

Abureema, S., Smooker, P., Malmo, J. and Deighton, M. (2014). Molecular epidemiology of recurrent clinical mastitis due to *Streptococcus uberis*: Evidence of both an environmental source and recurring infection with the same strain. *Journal of Dairy Science*, **97**, 285-290.

Albenzio, M., Taibi, L., Muscio, A. and Sevi, A. (2002). Prevalence and etiology of subclinical mastitis in intensively managed flocks and related changes in the yield and quality of ewe milk. *Small Ruminant Research*, **43**, 219-226.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **215**, 403-410.

Antiabong, J. F., Boardman, W., Smith, I., Brown, M. H., Ball, A. S. and Goodman, A. E. (2013). A molecular survey of a captive wallaby population for periodontopathogens and the co-incidence of *Fusobacterium necrophorum* subspecies *necrophorum* with periodontal diseases. *Veterinary Microbiology*, **163**, 335-343.

Baker, G. C., Smith, J. J. and Cowan, D. A. (2003). Review and re-analysis of domain-specific 16S primers. *Microbiology Methods*, **55**, 541-555.

Bala, M., Kumar, S., Raghava, G. P. S. and Mayilraj, S. (2013). Draft genome sequence of *Rhodococcus qingshengii* strain BKS 2-40. *Genome Announcements*, **1**, 128.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., Karbelashvili, M. S., Kirk, S. M., Li, H., Liu, X., Maisinger, K. S., Murray, L. J., Obradovic, B., Ost, T., Parkinson, M. L., Pratt, M. R., Rasolonjatovo, I. M. J., Reed, M. T., Rigatti, R., Rodighiero, C., Ross, M. T., Sabot, A., Sankar, S. V., Scally, A., Schroth, G. P., Smith, M. E., Smith, V. P., Spiridou, A., Torrance, P. E., Tzonev, S. S., Vermaas, E. H., Walter, K., Wu, X., Zhang, L., Alam, M. D., Anastasi, C., Aniebo, I. C., Bailey, D. M. D., Bancarz, I. R., Banerjee, S., Barbour, S. G., Baybayan, P. A., Benoit, V. A., Benson, K. F., Bevis, C., Black, P. J., Boodhun, A., Brennan, J. S., Bridgham, J. A., Brown, R. C., Brown, A. A., Buermann, D. H., Bundu, A. A., Burrows, J. C., Carter, N. P., Castillo, N., Chiara E. Catenazzi, M., Chang, S., Neil Cooley, R., Crake, N. R., Dada, O. O., Diakoumakos, K. D., Dominguez-Fernandez, B., Earnshaw, D. J., Egbujor, U. C., Elmore, D. W., Etchin, S. S., Ewan, M. R., Fedurco, M., Fraser, L. J., Fuentes Fajardo, K. V., Scott Furey, W., George, D., Gietzen, K. J., Goddard, C. P., Golda, G. S., Granieri, P. A., Green, D. E., Gustafson, D. L., Hansen, N. F., Harnish, K., Haudenschild, C. D., Heyer, N. I., Hims, M. M., Ho, J. T., Horgan, A.

M., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, **456**, 53-59.

Berriatua, E., Ziluaga, I., Miguel-Virto, C., Uribarren, P., Juste, R., Laevens, S., Vandamme, P. and Govan, J. R. W. (2001). Outbreak of subclinical mastitis in a flock of dairy sheep associated with *Burkholderia cepacia* complex infection. *Journal of Clinical Microbiology*, **39**, 990-994.

Bhatt, V. D., Ahir, V. B., Koringa, P. G., Jakhesara, S. J., Rank, D. N., Nauriyal, D. S., Kunjadia, A. P. and Joshi, C. G. (2012). Milk microbiome signatures of subclinical mastitis-affected cattle analysed by shotgun sequencing. *Journal of Applied Microbiology*, **112**, 639-650.

Biberstein, E. L., Shreeve, B. J. and Thompson, R. A. (1970). Variation in carrier rates of *Pasteurella haemolytica* in sheep flocks. *International Journal of Comparative Pathology*, **80**, 499-507.

Biffa, D., Debela, E. and Beyene, F. (2005). Prevalence and risk factors of mastitis in lactating dairy cows in southern Ethiopia. *International Journal of Applied Research in Veterinary Medicine*, **3**, 189-198.

Billion, P. and Decremoux, R. (1998). Mastitis of dairy ewes: etiology, detection and control. Dairy Sheep Symposium., Available at: http://bit.ly/YdGHG8. (Accessed 25-10-2009).

Bishop, H., Erkelens, J. and Van Winden, S. (2010). Predictors for successful bacteriological culture from milk samples. *Veterinary Record*, **166**, 322-324.

Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. L., Wertheim-van Dillen, P. M. E. and Noordaa, J. (1990). Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology*, **28**, 495-503.

Bradley, A. J. and Green, L. E. (2001a). Adaptation of *Escherichia coli* to the bovine mammary gland. *Journal of Clinical Microbiology*, **39**, 1845-1849.

Bradley, A. J. and Green, M. J. (2001b). Aetiology of clinical mastitis in six Somerset dairy herds. *Veterinary Record*, **148**, 683-686.

Bradley, A. J., Leach, K. A., Breen, J. E., Green, L. E. and Green, M. J. (2007). Survey of the incidence and aetiology of mastitis on dairy farms in England and Wales. *Veterinary Record*, **160**, 253-258.

Braem, G., De Vliegher, S., Verbist, B., Heyndrickx, M., Leroy, F. and De Vuyst, L. (2012). Culture-independent exploration of the teat apex microbiota of dairy cows reveals a wide bacterial species diversity. *Veterinary Microbiology*, **157**, 383-390.

Cabrera-Rubio, R., Collado, M., Laitinen, K., Salminen, S., Isolauri, E. and Mira, A. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *American Society for Nutrition*, **96**, 544-551.

Calvo-Bado, L. A., Oakley, B. B., Dowd, S. E., Green, L. E., Medley, G. F., Ul-Hassan, A., Bateman, V., Gaze, W., Witcomb, L., Grogono-Thomas, R., Kaler, J., Russell, C. L. and Wellington, E. M. H. (2011). Ovine pedomics: the first study of the ovine foot 16S rRNA-based microbiome. *ISME Journal*, **5**, 1426-1437.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Tumbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335-336.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G. and Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, **6**, 1621-1624.

Chen, H.-C. and Hwang, W.-Z. (2008). Development of a PCR-denaturing gradient gel electrophoresis method targeting the tuf gene to differentiate and identify *Staphylococcus* species. *Journal of Food and Drug Analysis*, **16**, 44-51.

Cho, I. and Blaser, M. J. (2012). The human microbiome: at the interface of health and disease. *Nature Reviews Genetics*, **13**, 260-270.

Chui, L. W., King, R., Lu, P., Manninen, K. and Sim, J. (2004). Evaluation of four DNA extraction methods for the detection of *Mycobacterium avium* subsp. *paratuberculosis* by polymerase chain reaction. *Diagnostic Microbiology and Infectious Disease*, **48**, 39-45.

Clarke, K. R. and Gorley, R. N. (2006). *PRIMER V6: User Manual/Tutorial.* **In**: PRIMER-E, Plymouth.

Contreras, A., Sierra, D., Sanchez, A., Corrales, J. C., Marco, J. C., Paape, M. J. and Gonzalo, C. (2007). Mastitis in small ruminants. *Small Ruminant Research*, **68**, 145-153.

Coppola, S., Blaiotta, G., Ercolini, D. and Moschetti, G. (2001). Molecular evaluation of microbial diversity occurring in different types of Mozzarella cheese. *Journal of Applied Microbiology*, **90**, 414-420.

Costello, E., Lauber, C., Hamady, M., Fierer, N., Gordon, J. and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science*, **326**, 1694-1697.

Cremonesi, P., Castiglioni, B., Malferrari, G., Biunno, G., Vimercati, C., Moroni, P., Morandi, S. and Luzzana, M. (2006). Technical note: Improved method for rapid DNA extraction of mastitis pathogens directly from milk. *Journal of Dairy Science*, **89**, 163-169.

Cullen, G. A. (1966). The ecology of *Streptococcus uberis*. *British Veterinary Journal*, **122**, 333-336.

Daly, P., Collier, T. and Doyle, S. (2002). PCR-ELISA detection of *Escherichia coli* in milk. *Letters in Applied Microbiology*, **34**, 222-226.

Davey, M. E. and O'Toole, G. A. (2000). Microbial biofilms: from ecology to molecular genetics. *Microbiology and Molecular Biology Reviews*, **64**, 847-867.

De Baets, F., Schelstraete, P., Haerynck, F., Van Biervliet, S., De Bruyne, R., Franckx, H. and Van Daele, S. (2014). *Achromobacter xylosoxidans* induced bronchiolitis obliterans in cystic fibrosis. *Pediatric Pulmonology*, **49**, 414-416.

Dekio, I., Hayashi, H., Sakamoto, M., Kitahara, M., Nishikawa, T., Suematsu, M. and Benno, Y. (2005). Detection of potentially novel bacterial components of the human skin microbiota using culture-independent molecular profiling. *Journal of Medical Microbiology*, **54**, 1231-1238.

Delbes, C., Ali-Mandjee, L. and Montel, M.C. (2007). Monitoring bacterial communities in raw milk and cheese by culture-dependent and independent 16S rRNA gene-based analyses. *Applied and Environmental Microbiology*, **73**, 1882-1891.

Delgado, S., Arroyo, R., Martin, R. and Rodriguez, J. M. (2008). PCR-DGGE assessment of the bacterial diversity of breast milk in women with lactational infectious mastitis. *BMC Infectious Diseases*, **8**, 51-58.

Delgado, S., Rachid, C. T. C. C., Fernandez, E., Rychlik, T., Alegria, A., Peixoto, R. S. and Mayo, B. (2013). Diversity of thermophilic bacteria in raw, pasteurized and selectively-cultured milk, as assessed by culturing, PCR-DGGE and pyrosequencing. *International Journal of Food Microbiology*, **36**, 103-111.

Deluyker, H. A., Gay, J. M. and Weaver, L. D. (1993). Interrelationships of somatic cell count, mastitis, and milk yield in a low somatic cell count herd [ABSTRACT ONLY]. *Journal of Dairy Science*, **76**, 3445-3452.

Dohoo, I. R. and Meek, A. H. (1982). Somatic cell counts in bovine milk. *Canadian Veterinary Journal*, **23**, 119-125.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460-2461.

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996-998.

Edwards, U., Rogall, T., Blocker, H., Emde, M. and Bottger, E. (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research*, **17**, 7843-7853.

El-Khodery, S. A. and Osman, S. A. (2008). Acute coliform mastitis in buffaloes (Bubalus bubalis): Clinical findings and treatment outcomes. *Tropical Animal Health and Production*, **40**, 93-99.

Embley, T. M. (1991). The linear PCR reaction: a simple and robust method for sequencing amplified rRNA genes. *Letters in Applied Microbiology*, **13**, 171-174.

Ercolini, D. (2004). PCR-DGGE fingerprinting: novel strategies for detection of microbes in food. *Microbiology Methods*, **56**, 297-314.

Erlwein, O., Robinson, M. J., Dustan, S., Weber, J., Kaye, S. and McClure, M. O. (2011). DNA extraction columns contaminated with murine sequences. *PLoS ONE*, 6, Available at: 10.1371/journal.pone.0023484. (Accessed 03/09/14).

Evans, G. E., Murdoch, D. R., Anderson, T. P., Potter, H. C., George, P. M. and Chambers, S. T. (2003). Contamination of Qiagen DNA extraction kits with Legionella DNA. *Journal of Clinical Microbiology*, **41**, 3452-3453.

Extramiana, A. B., González, L., Cortabarría, N., García, M. and Juste, R. A. (2002). Evaluation of a PCR technique for the detection of Maedi-Visna proviral DNA in blood, milk and tissue samples of naturally infected sheep. *Small Ruminant Research*, **44**, 109-118.

Farell, E. M. and Alexandre, G. (2012). Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. *BMC research notes*, **5**, 257-257.

Fernández, L., Langa, S., Martín, V., Maldonado, A., Jiménez, E., Martín, R. and Rodríguez, J. M. (2012). The human milk microbiota: Origin and potential roles in health and disease. *Pharmacological Research*, **69**, 1-10.

Ferris, M. J., Muyzer, G. and Ward, D. M. (1996). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Applied and Environmental Microbiology*, **62**, 340-346.

Flórez, A. B. and Mayo, B. (2006). Microbial diversity and succession during the manufacture and ripening of traditional, Spanish, blue-veined Cabrales cheese, as determined by PCR-DGGE. *Journal of Food Microbiology*, **110**, 165-171.

Fthenakis, G. C. (1996). Somatic cell counts in milk of Welsh-Mountain, Dorset-Horn and Chios ewes throughout lactation. *Small Ruminant Research*, **20**, 155-162.

Fthenakis, G. C. and Jones, J. E. T. (1990). The effect of experimentally induced subclincal mastitis on milk-yield of ewes and growth of lambs. *British Veterinary Journal*, **146**, 43-49.

Fudou, R., Jojima, Y., Seto, A., Yamada, K., Kimura, E., Nakamatsu, T., Hiraishi, A. and Yamanaka, S. (2002). *Corynebacterium efficiens* sp nov., a glutamic-acid-producing species from soil and vegetables. *International Journal of Systematic and Evolutionary Microbiology*, **52**, 1127-1131.

Giannino, M. L., Marzotto, M., Dellaglio, F. and Feligini, M. (2009). Study of microbial diversity in raw milk and fresh curd used for Fontina cheese production by culture-independent methods. *International Journal of Food Microbiology*, **130**, 188-195.

Gini, G. A. (1990). Ocular infection cause by *Psychrobacter immobilis* acquired in the hospital. *Journal of Clinical Microbiology*, **28**, 400-401.

Giraffa, G. and Neviani, E. (2001). DNA-based, culture-independent strategies for evaluating microbial communities in food-associated ecosystems. *International Journal of Food Microbiology*, **67**, 19-34.

Gonzalez-Rodriguez., Gonzalo, C., Primitivo, F. S. and Carmenes, P. (1995). Relationship between somatic cell count and intramammary infection of the half udder in dairy ewes. *Journal of Dairy Science*, **78**, 2753-2759.

Govan, J. R. W., Hughes, J. E. and Vandamme, P. (1996). *Burkholderia cepacia*: Medical, taxonomic and ecological issues. *Journal of Medical Microbiology*, **45**, 395-407.

Green, M. J., Bradley, A. J., Medley, G. F. and Browne, W. J. (2007). Cow, farm and management factors during the dry period that determine the rate of clinical mastitis after calving. *Journal of Dairy Science*, **90**, 3764-3776.

Green, M. J., Bradley, A. J., Newton, H. and Browne, W. J. (2006). Seasonal variation of bulk milk somatic cell counts in UK dairy herds: Investigations of the summer rise. *Preventative Veterinary Medicine*, **74**, 293-308.

Green, M. J., Cripps, P. J. and Green, L. E. (1998). Prognostic indicators for toxic mastitis in dairy cows. *Veterinary Record*, **143**, 127-130.

Green, M. J., Green, L. E., Bradley, A. J., Burton, P. R., Schukken, Y. H. and Medley, G. F. (2005). Prevalence and associations between bacterial isolates from dry mammary glands of dairy cows. *Veterinary Record*, **156**, 71-77.

Green, M. J., Green, L. E. and Cripps, P. J. (1997). Comparison of fluid and flunixin meglumine therapy in combination and individually in the treatment of toxic mastitis. *Veterinary Record*, **140**, 149-152.

Green, M. J., Green, L. E., Medley, G. F., Schukken, Y. H. and Bradley, A. J. (2002). Influence of dry period bacterial intramammary infection on clinical mastitis in dairy cows. *Journal of Dairy Science*, **85**, 2589-2599.

Grice, E. A., Kong, H. H., Renaud, G., Young, A. C., Bouffard, G. G., Blakesley, R. W., Wolfsberg, T. G., Turner, M. L., Segre, J. A. and Nisc Comparative, S. (2008). A diversity profile of the human skin microbiota. *Genome Research*, **18**, 1043-1050.

Grohn, Y. T., Wilson, D. J., Gonzalez, R. N., Hertl, J. A., Schulte, H., Bennett, G. and Schukken, Y. H. (2004). Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. *Journal of Dairy Science*, **87**, 3358-3374.

Guo, X. Q., Li, R., Zheng, L. Q., Lin, D. Q., Sun, J. Q., Li, S. P., Li, W. J. and Jiang, J. D. (2010). *Jeotgalicoccus huakuii* sp nov., a halotolerant bacterium isolated from seaside soil. *International Journal of Systematic and Evolutionary Microbiology*, **60**, 1307-1310.

Harmon, R. J. (1994). Physiology of mastitis and factors affecting somatic cell counts. *Journal of Dairy Science*, **77**, 2103-2112.

Heras, A. L., Vela, A. I., Fernandez, E., Legaz, E., Dominguez, L. and Fernandez-Garayzabal, J. F. (2002). Unusual outbreak of clinical mastitis in dairy sheep caused by *Streptococcus equi* subsp. *zooepidemicus*. *Clinical Microbiology*, **40**, 1106-1108.

Hillerton, J. E. and Berry, E. A. (2005). Treating mastitis in the cow- a tradition or an archaism. *Journal of Applied Microbiology*, **98**, 1250-1255.

Hogan, J. S., Gonzalez, R. W., Harmon, R. S., Nickerson, S. C., Oliver, S. P., Pankey, J. W. and Smith, K. L. (1999). Laboratory handbook on bovine mastitis. (Madison, Wisconsin, USA: National Mastitis Council).

Holzapfel, W. H., Haberer, P., Snel, J., Schillinger, U. and Huis in't Veld, J. H. J. (1998). Overview of gut flora and probiotics. *International Journal of Food Microbiology*, **41**, 85-101.

Hoshino, Y., Mukai, A., Yazawa, K., Uno, J., Ishikawa, J., Ando, A., Fukai, T. and Mikami, Y. (2004). Transvalencin A, a thiazolidine zinc complex antibiotic produced by a clinical isolate of *Nocardia transvalensis*. *The Journal of antibiotics*, **57**, 797-802.

Hue, S., Gray, E. R., Gall, A., Katzourakis, A., Tan, C. P., Houldcroft, C. J., McLaren, S., Pillay, D., Futreal, A., Garson, J. A., Pybus, O. G., Kellam, P. and Towers, G. J. (2010). Disease-associated XMRV sequences are consistent with laboratory contamination. *Retrovirology*, **7**, 111-121.

Hunt, K. M., Foster, J. A., Forney, L. J., Schütte, U. M. E., Beck, D. L., Abdo, Z., Fox, L. K., Williams, J. E., McGuire, M. K. and McGuire, M. A. (2011). Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS ONE*, 6, Available at: 10.1371/journal.pone.0021313. (Accessed 27/08/2011).

Huntley, S. J., Cooper, S., Bradley, A. J. and Green, L. E. (2012). A cohort study of the associations between udder conformation, milk somatic cell count, and lamb weight in suckler ewes. *Journal of Dairy Science*, **95**, 5001-5010.

Jara, S., Sanchez, M., Vera, R., Cofre, J. and Castro, E. (2011). The inhibitory activity of *Lactobacillus* spp. isolated from breast milk on gastrointestinal pathogenic bacteria of nosocomial origin. *Anaerobe*, **17**, 474-477.

Jones, J. E. T. (1985). *Conference Proceedings: An investigation into mastitis in sheep: preliminary phase*. UK, **10**, 48-51.

Jones, J. E. T. and Watkins, G. H. (2000). Mastitis and contagious agalactiae. In *Diseases of sheep*, 3 edn, pp. 99-105. Edited by W. B. Martin and I. D. Aitken: Blackwell Oxford.

Jost, B. H. and Billington, S. J. (2005). *Arcanobacterium pyogenes*: molecular pathogenesis of an animal opportunist. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, **88**, 87-102.

Jousimies-Somer, H., Pyorala, S. and Kanervo, A. (1996). Susceptibilities of bovine summer mastitis bacteria to antimicrobial agents. *Antimicrobial Agents and Chemotherapy*, **40**, 157-60.

Junemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., Mellmann, A., Goesmann, A., von Haeseler, A., Stoye, J. and Harmsen, D. (2013). Updating benchtop sequencing performance comparison. *Nature Biotechnology*, **31**, 294-296.

Kampfer, P., Albrecht, A., Buczolits, S. and Busse, H. J. (2002). *Psychrobacter faecalis* sp nov., a new species from a bioaerosol originating from pigeon faeces. *Systematic and Applied Microbiology*, **25**, 31-36.

Keisler, D. H., Andrews, M. L. and Moffatt, R. J. (1992). Subclinical mastitis in ewes and its effects in lamb performance. *Animal Science*, **70**, 1677-1681.

Knox, K., Carrigan, D., Simmons, G., Teque, F., Zhou, Y. C., Hackett, J., Qiu, X. X., Luk, K. C., Schochetman, G., Knox, A., Kogelnik, A. M. and Levy, J. A. (2011). No evidence of Murine-like Gammaretroviruses in CFS patients previously identified as XMRV-infected. *Science*, **333**, 94-97.

Koskinen, M. T., Wellenberg, G. J., Sampimon, O. C., Holopainen, J., Rothkamp, A., Salmikivi, L., van Haeringen, W. A., Lam, T. and Pyorala, S. (2010). Field comparison of real-time polymerase chain reaction and bacterial culture for identification of bovine mastitis bacteria. *Journal of Dairy Science*, **93**, 5707-5715.

Kuang, Y., Tani, K., Synnott, A. J., Ohshima, K., Higuchi, H., Nagahata, H. and Tanji, Y. (2009). Characterization of bacterial population of raw milk from bovine mastitis by culture-independent PCR–DGGE method. *Biochemical Engineering Journal*, **45**, 76-81.

Kuehn, J. S., Gorden, P. J., Munro, D., Rong, R. C., Dong, Q. F., Plummer, P. J., Wang, C. and Phillips, G. J. (2013). Bacterial community profiling of milk samples as a means to understand culture-negative bovine clinical mastitis. *PLoS ONE*, **8**, Available at: 10.1371/journal.pone.0061959. (Accessed 15/04/14).

Lafarge, V., Ogier, J. C., Girard, V., Maladen, V., Leveau, J. Y., Gruss, A. and Delacroix-Buchet, A. (2004). Raw cow milk bacterial population shifts attributable to refrigeration. *Applied and Environmental Microbiology*, **70**, 5644-5650.

Lafi, S. Q. (2006). Use of somatic cell counts and California mastitis test results from udder halves milk samples to detect subclinical intramammary infection in Awassi sheep. *Small Ruminant Research*, **62**, 83-86.

Lechtenberg, K. F., Nagaraja, T. G. and Chengappa, M. M. (1998). Antimicrobial susceptibility of *Fusobacterium necrophorum* isolated from bovine hepatic abscesses. *American Journal of Veterinary Research*, **59**, 44-47.

Lechtenberg, K. F., Nagaraja, T. G., Leipold, H. W. and Chengappa, M. M. (1988). Bacteriologic and histologic-studies of hepatic-abscesses in cattle. *American Journal of Veterinary Research*, **49**, 58-62.

Leff, L., Dana, J., McArthur, J. and Shimkets, L. (1998). Comparison of methods of DNA extraction from stream sediments. *Applied Environmental Microbiology*, **61**, 1141-1143.

Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658-1659.

Liu, Z. Z., Lozupone, C., Hamady, M., Bushman, F. D. and Knight, R. (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research*, **35**. Available at: 10.1093/nar/gkm541. (Accessed 02/09/14).

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. and Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, **30**, 434-440.

Lozano, F., Florez, C., Recio, F. J., Gamboa, F., Gomezmateos, J. M. and Martin, E. (1994). Fatal *Psychrobacter immobilis* infection in a patient with AIDS. *Aids*, **8**, 1189-1190.

Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J., Bachleitner, M. and Schleifer, K. H. (1998). Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis*, **19**, 554-568.

Luo, P., Hu, C., Zhang, L., Ren, C. and Shen, Q. (2007). Effects of DNA extraction and universal primers on 16S rRNA gene-based DGGE analysis of a bacterial community from fish farming water. *Chinese Journal of Oceanology and Limnology*, **25**, 310-316.

Macmillan, K. L., Duirs, G. F. and Duganzich, D. M. (1983). Association between dry cow therapy, clinical mastitis, and somatic cell count score with milk and fat production in ten New Zealand dairy herds. *Journal of Dairy Science*, **66**, 259-265.

Madsen, M., Aalbaek, B. and Hansen, J. W. (1992). Comparative bacteriological studies on summer mastitis in grazing cattle and pyogenes mastitis in stabled cattle in Denmark. *Veterinary Microbiology*, **32**, 81-88.

Mahenthiralingam, E., Baldwin, A. and Dowson, C. G. (2008). *Burkholderia cepacia* complex bacteria: opportunistic pathogens with important natural biology. *Journal of Applied Microbiology*, **104**, 1539-1551.

Malinowski, E., Lassa, H., Klossowska, A., Markiewicz, H., Kaczmarowski, M. and Smulski, S. (2006). Relationship between mastitis agents and somatic cell count in foremilk samples. *Bulletin of the Veterinary Institute in Pulawy*, **50**, 349-352.

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *Journal of Molecular Biology*, **3**, 208-218.

Marogna, G., Rolesu, S., Lollai, S., Tola, S. and Leori, G. (2010). Clinical findings in sheep farms affected by recurrent bacterial mastitis. *Small Ruminant Research*, **88**, 119-125.

Marsh, P. D. (1995). The role of microbiology in dental caries. *Advances in Dental Research.*, **9**, 244-254.

Martin, E., Klug, K., Frischmann, A., Busse, H. J., Kampfer, P. and Jackel, U. (2011). *Jeotgalicoccus coquinae* sp. nov. and *Jeotgalicoccus aerolatu*s sp. nov., isolated from poultry houses. *International Journal of Systematic and Evolutionary Microbiology*, **61**, 237-241.

Maruyama, A., Honda, D., Yamamoto, H., Kitamura, K. and Higashihara, T. (2000). Phylogenetic analysis of psychrophilic bacteria isolated from the Japan Trench, including a description of the deep-sea species *Psychrobacter pacificensis* sp nov. *International Journal of Systematic and Evolutionary Microbiology*, **50**, 835-846.

Mavrogenis, A. P., Koumas, A., Kakoyiannis, C. K. and Taliotis, E. C. (1996). Use of somatic cell count for detection of subclinical mastitis in sheep. *Somatic Cells and Milk of Small Ruminants*, 323-325.

Mavrogianni, V. S., Cripps, P. J. and Fthenakis, G. C. (2007). Bacterial flora and risk of infection of the ovine teat duct and mammary gland throughout lactation. *Preventive Veterinary Medicine*, **79**, 163-173.

McGillivery, D. J., Nicholls, T. J. and Hatch, P. H. (1984). Isolation of *Fusobacterium necrophorum* from a case of bovine mastitis. *Australian Veterinary Journal*, **61**, 325-325.

Meiri-Bendek, I., Lipkin, E., Friedman, A., Leitner, G., Saran, A., Friedman, S. and Kashi, Y. (2002). A PCR-based method for the detection of *Streptococcus agalactiae* in milk. *Journal of Dairy Science*, **85**, 1717-1723.

Methe, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., Gevers, D., Petrosino, J. F., Abubucker, S., Badger, J. H., Chinwalla, A. T., Earl, A. M., FitzGerald, M. G., Fulton, R. S., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., Versalovic, J., Wollam, A. M., Worley, K. C., Wortman, J. R., Young, S. K., Zeng, Q., Aagaard, K. M., Abolude, O. O., Allen-Vercoe, E., Alm, E. J., Alvarado, L., Andersen, G. L., Anderson, S., Appelbaum, E., Arachchi, H. M., Armitage, G., Arze, C. A., Ayvaz, T., Baker, C. C., Begg, L., Belachew, T., Bhonagiri, V., Bihan, M., Blaser, M. J., Bloom, T., Bonazzi, V. R., Brooks, P., Buck, G., Buhay, C. J., Busam, D. A., Campbell, J. L., Canon, S. R., Cantarel, B. L., Chain, P. S., Chen, I. M. A., Chen, L., Chhibba, S., Chu, K., Ciulla, D. M., Clemente, J. C., Clifton, S. W., Conlan, S., Crabtree, J., Cutting, M. A., Davidovics, N. J., Davis, C. C., DeSantis, T. Z., Deal, C., Delehaunty, K. D., Dewhisrst, F. E., Deych, E., Ding, Y., Dooling, D. J., Dugan, S. P., Dunne, W. M., Durkin, A. S., Edgar, R. C., Erlich, R. L., Farmer, C. N., Farrell, R. M., Faust, K., Feldgarden, M., Felix, V. M., Fisher, S., Fodor, A. A., Forney, L., Foster, L., Di Francesco, V., Friedman, J., Friedrich, D. C., Fronick, C. C., Fulton, L. L., Gao, H., Garcia, N., Giannoukos, G., Giblin, C., Giovanni, M. Y., Goldberg, J. M., et al. (2012). A framework for human microbiome research. *Nature*, **486**, 215-221.

Millar, B. C., Jiru, X., Moore, J. E. and Earle, J. A. P. (2000). A simple and sensitive method to extract bacterial, yeast and fungal DNA from blood culture material. *Journal of Microbiological Methods*, **42**, 139-147.

Monaghan, E. (2010). *Development and use of a multiplex PCR to detect common mastitis pathogens in ewes' milk*. **In**: Department of Biological Sciences. University of Warwick.

Morgan, X. C., Segata, N. and Huttenhower, C. (2013). Biodiversity and functional genomics in the human microbiome. *Trends in Genetics*, **29**, 51-58.

Morita, H., Shiratori, C., Murakami, M., Takami, H., Toh, H., Kato, Y., Nakajima, F., Takagi, M., Akita, H., Masaoka, T. and Hattori, M. (2008). *Sharpea azabuensis* gen. nov., sp nov., a Gram-positive, strictly anaerobic bacterium isolated from the faeces of thoroughbred horses. *International Journal of Systematic and Evolutionary Microbiology*, **58**, 2682-2686.

Moschetti, G., Blaiotta, G., Aponte, M., Mauriello, G., Villani, F. and Coppola, S. (1997). Genotyping of *Lactobacillus delbrueckii* subsp. bulgaricus and determination of the number and forms of rrn operons in *L. delbrueckii* and its subspecies. *Research in Microbiology*, **148**, 501-510.

Muyzer, G. (1999). DGGE/TGGE a method for identifying genes from natural ecosystems. *Current Opinion in Microbiology*, **2**, 317-322.

Muyzer, G., Dewaal, E. C. and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes coding for 16S ribosomal RNA. *Applied and Environmental Microbiology*, **59**.

Muyzer, G. and Schäfer, H. (2001). Denaturing gradient gel electrophoresis in marine microbial ecology. *Methods in Microbiology*, **30**, 425-468.

Muyzer, G. and Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, **73**, 127-141.

Naccache, S. N., Greninger, A. L., Lee, D., Coffey, L. L., Phan, T., Rein-Weston, A., Aronsohn, A., Hackett, J., Delwart, E. L. and Chiu, C. Y. (2013). The perils of pathogen discovery: Origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *Journal of Virology*, **87**, 11966-11977.

Nagaraja, T. G., Narayanan, S. K., Stewart, G. C. and Chengappa, M. M. (2005). *Fusobacterium necrophorum* infections in animals: Pathogenesis and pathogenic mechanisms. *Anaerobe*, **11**, 239-246.

Nath, S. R., Mathew, A. P., Mohan, A. and Anila, K. R. (2013). *Rhodococcus equi* granulomatous mastitis in an immunocompetent patient. *Journal of Medical Microbiology*, **62**, 1253-1255.

Nubel, U., Engelen, B., Felske, A., Snaidr, J., Wieshuber, A., Amann, R. I., Ludwig, W. and Backhaus, H. (1996). Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *Journal of Bacteriology*, **178**, 5636-5643.

O' Reilly, K. M., Green, M. J., Peeler, E. J., Fitzpatrick, J. L. and Green, L. E. (2006). Investigation of risk factors for clinical mastitis in British dairy herds with bulk milk somatic cell counts less than 150,000 cells/ml. *Veterinary Record*, **158**, 649-653.

Oakes, B., Tai, A., Cingöz, O., Henefield, M., Levine, S., Coffin, J. and Huber, B. (2010). Contamination of human DNA samples with mouse DNA can lead to false detection of XMRV-like sequences. *Retrovirology*, **7**, 1-10.

Odumeru, J., Gao, A., Chen, S., Raymond, M. and Mutheria, L. (2001). Use of bead beater for preparation of *Mycobacterium paratuberculosis* template DNA in milk. *Canadian Veterinary Research*, **65**, 201-205.

Ogier, J. C., Son, O., Gruss, A., Tailliez, P. and Delacroix-Buchet, A. (2002). Identification of the bacterial microflora in dairy products by temporal temperature gradient gel electrophoresis. *Applied and Environmental Microbiology*, **68**, 3691-3701.

Ogino, A., Koshikawa, H., Nakahara, T. and Uchiyama, H. (2001). Succession of microbial communities during a biostimulation process as evaluated by DGGE and clone library analyses. *Journal of Applied Microbiology*, **91**, 625-635.

Oikonomou, G., Bicalho, M. L., Meira, E., Rossi, R. E., Foditsch, C., Machado, V. S., Teixeira, A. G. V., Santisteban, C., Schukken, Y. H. and Bicalho, R. C. (2014). Microbiota of cow's milk; distinguishing healthy, sub-clinically and clinically diseased quarters. *PLoS ONE*, **9**, Available at: 10.1371/journal.pone.0085904. (Accessed 01/02/14).

Oikonomou, G., Machado, V. S., Santisteban, C., Schukken, Y. H. and Bicalho, R. C. (2012). Microbial diversity of bovine mastitic milk as described by pyrosequencing of metagenomic 16S rDNA. *PLoS ONE*, **7**, Available at: 10.1371/journal.pone.0047671. (Accessed 25/06/13).

Oikonomou, G., Teixeira, A. G. V., Foditsch, C., Bicalho, M. L., Machado, V. S. and Bicalho, R. C. (2013). Fecal microbial diversity in pre-weaned dairy calves as described by pyrosequencing of metagenomic 16S rDNA. Associations of *Faecalibacterium* species with health and growth. *PLoS ONE*, **8**, Available at: 10.1371/journal.pone.0063157. (Accessed 02/07/14).

Oliver, S. P., Gillespie, B. E. and Jayarao, B. M. (1998). Detection of new and persistent *Streptococcus uberis* and *Streptocococcus dysgalactiae* intramammary infections by polymerase chain reaction-based DNA fingerprinting. *FEMS Microbiology Letters*, **160**, 69-73.

Omaleki, L., Browning, G. F., Allen, J. L. and Barber, S. R. (2011). The role of Mannheimia species in ovine mastitis. *Veterinary Microbiology*, **153**, 67-72.

Onnasch, H., Healy, A. M., Brophy, P., Kinsella, A. and Doherty, M. L. (2002). A study of mastitis in Irish sheep. *Research in Veterinary Science*, **72**, 42.

Oros-Sichler, M., Gomes, N. C. M., Neuber, G. and Smalla, K. (2006). A new semi-nested PCR protocol to amplify large 18S rRNA gene fragments for PCR-DGGE analysis of soil fungal communities. *Journal of Microbiological Methods*, **65**, 63-75.

Ovreas, L., Forney, L., Daae, F. L. and Torsvik, V. (1997). Distribution of bacterioplankton in meromictic Lake Saelenvannet, as determined by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. *Applied and Environmental Microbiology*, **63**, 3367-3373.

Panciera, R. J., Perino, L. J., Baldwin, C. A., Morton, R. J. and Swanson, J. E. (1989). Observations on calf diphtheria in the commercial feedlot. *Agri-Practice*, **10**, 12-17.

Pankey, J. W., Nickerson, S. C., Boddie, R. L. and Hogan, J. S. (1985). Effects of *Corynebacterium bovis* infection on susceptibility to major mastitis pathogens. *Journal of Dairy Science*, **68**, 2684-2693.

Park, Y. K., Koo, H. C., Kim, S. H., Hwang, S. Y., Jung, W. K., Kim, J. M., Shin, S., Kim, R. T. and Park, Y. H. (2007a). The analysis of milk components and pathogenic bacteria isolated from bovine raw milk in Korea. *Journal of Dairy Science*, **90**, 5405-5414.

Park, Y. W., Juarez, M., Ramos, M. and Haenlein, G. F. W. (2007b). Physio-chemical characteristics of goat and sheep milk. *Small Ruminant Research*, **68**, 88-113.

Peeler, E. J., Green, M. J. and Fitzpatrick, J. L. (2002). Study of clinical mastitis in British dairy herds with bulk milk somatic cell counts less than 150,000 cells/ml. *Veterinary Record*, **151**, 170-176.

Pengov, A. (2001). The role of coagulase-negative *staphylococcus* spp. and associated somatic cell counts in the ovine mammary gland. *Journal of Dairy Science*, **84**, 572-574.

Phuektes, P., Mansell, P. D., Dyson, R. S., Hooper, N. D., Dick, J. S. and Browning, G. F. (2001). Molecular epidemiology of *Streptococcus uberis* isolates from dairy cows with mastitis. *Journal of Clinical Microbiology*, **39**, 1460-1466.

Prieto, M., Garciaarmesto, M. R., Garcialopez, M. L., Otero, A. and Moreno, B. (1992). Numerical taxonomy of Gram-negative, nonmotile, nonfermentative bacteria isolated during chilled storage of lamb carcasses. *Applied and Environmental Microbiology*, **58**, 2245-2249.

Proctor, Lita M. (2011). The Human Microbiome Project in 2011 and Beyond. *Cell host & microbe*, **10**, 287-291.

Psifidi, A., Dovas, C. I. and Banos, G. (2010). A comparison of six methods for genomic DNA extraction suitable for PCR-based genotyping applications using ovine milk samples. *Molecular and Cellular Probes*, **24**, 93-98.

Purdy, K. J. (2005). Nucleic acid recovery from complex environmental samples. *Methods in Enzymology*, **397**, 271-292.

Purdy, K. J. (1996). *The use of 16S rRNA-targeted oligonucleotide probes to study the ecology of sulphate-reducing bacteria*. **In**: Department of Biological and Chemical Sciences. University of Essex.

Purdy, K. J., Nedwell, D. B., Embley, T. H. and Takii, S. (1997). Use of 16S rRNA-targeted oligonucleotide probes to investigate the occurrence and selection of sulphate-reducing bacteria in response to nutrient addition to sediment slurry microcosms from a Japanese estuary. *FEMS Microbiology Ecology*, **24**, 231-234.

Pyorala, S., Jousimiessomer, H. and Mero, M. (1992). Clinical, bacteriological and therapeutic aspects of bovine mastitis caused by aerobic and anaerobic pathogens. *British Veterinary Journal*, **148**, 54-62.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. and Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341-353.

Quigley, L., O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F. and Cotter, P. D. (2012). A comparison of methods used to extract bacterial DNA from raw milk and raw milk cheese. *Journal of Applied Microbiology*, **113**, 96-105.

Quigley, L., O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F. and Cotter, P. D. (2011). Molecular approaches to analysing the microbial composition of raw milk and raw milk cheese. *International Journal of Food Microbiology*, **150**, 81-94.

Raats, D., Offek, M., Minz, D. and Halpern, M. (2011). Molecular analysis of bacterial communities in raw cow milk and the impact of refrigeration on its structure and dynamics. *International Journal of Food Microbiology*, **28**, 465-471.

Ramesh, A., Padmapriya, B. P., Chandrashekar, A. and Varadaraj, M. C. (2002). Application of a convienent DNA extraction method and multiplex PCR for the direct detection of *Staphylococcus aureus* and *Yersinia enterocolitica* in milk samples. *Molecular and Cellular Probes*, **16**, 307-314.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M. and Cameron, B. (2009). *MLwIN version 2.1*. **In**: Centre for Multilevel Modelling, University of Bristol.

Rasolofo, E. A., St-Gelais, D., LaPointe, G. and Roy, D. (2010). Molecular analysis of bacterial population structure and dynamics during cold storage of untreated and treated milk. *International Journal of Food Microbiology*, **138**, 108-118.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G. M., Koenig, S. S. K., McCulle, S. L., Karlebach, S., Gorle, R., Russell, J., Tacket, C. O., Brotman, R. M., Davis, C. C., Ault, K., Peralta, L. and Forney, L. J. (2011). Vaginal microbiome of reproductive-age women. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 4680-4687.

Reneau, J. K. (1986). Effective use of dairy herd improvement somatic cell counts in mastitis control. *Journal of Dairy Science*, **69**, 1708-1720.

Reysenbach, A. L., Giver, L. J., Wickham, G. S. and Pace, N. R. (1992). Differential amplification of ribosomal-RNA genes by polymerase chain reaction. *Applied and Environmental Microbiology*, **58**, 3417-3418.

Reysenbach, A. L., Wickham, G. S. and Pace, N. R. (1994). Phylogenetic analysis of the hyperthermophilic pink filament in community octopus spring, Yellowstone National Park. *Applied and Environmental Microbiology*, **60**, 2113-2119.

Riffon, R., Sayasith, K., Khalil, H., Dubreuil, P., Drolet, M. and Lagace, J. (2001). Development of a rapid and sensitive test for the detection of major pathogens in bovine mastitis by PCR. *Journal of Clinical Microbiology*, **39**, 2584-2589.

Robinson, M., Erlwein, O., Kaye, S., Weber, J., Cingoz, O., Patel, A., Walker, M., Kim, W.-J., Uiprasertkul, M., Coffin, J. and McClure, M. (2010). Mouse DNA contamination in human tissue tested for XMRV. *Retrovirology*, **7**, 1-6.

Romanenko, L. A., Lysenko, A. M., Rohde, M., Mikhailov, V. V. and Stackebrandt, E. (2004). *Psychrobacter maritimus* sp nov and *Psychrobacter arenosus* sp nov., isolated from coastal sea ice and sediments of the Sea of Japan. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 1741-1745.

Rossello-Mora, R., Thamdrup, B., Schafer, H., Weller, R. and Amann, R. (1999). The response of the microbial community of marine sediments to organic carbon input under anaerobic conditions. *Systematic and Applied Microbiology*, **22**, 237-248.

Rossen, L., Nørskov, P., Holmstrøm, K. and Rasmussen, O. F. (1992). Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. *International Journal of Food Microbiology*, **17**, 37-45.

Rowe, H. A., Poxton, I. R. and Donachie, W. (2001). Survival of *Mannheimia (Pasteurella) haemolytica* in tracheobronchial washings of sheep and cattle. *Veterinary Microbiology*, **81**, 305-314.

Sato, E., Furuta, R. and Miyazawa, T. (2010). An endogenous murine leukemia viral genome contaminant in a commercial RT-PCR Kit is amplified using standard primers for XMRV. *Retrovirology*, **7**, 1-7.

Schukken, Y. H., Wilson, D. J., Welcome, F., Garrison-Tikofsky, L. and Gonzalez, R. N. (2003). Monitoring udder health and milk quality using somatic cell counts. *Veterinary Research*, **34**, 579-596.

Scott, M. J. and Jones, J. E. T. (1998). The carriage of *Pasteurella haemolytica* in sheep and its transfer between ewes and lambs in relation to mastitis. *Journal of Comparative Pathology*, **118**, 359-363.

Shu, M. Y., Wang, Y. H., Yu, J. H., Kuo, S., Coda, A., Jiang, Y., Gallo, R. L. and Huang, C. M. (2013). Fermentation of *Propionibacterium acnes*, a commensal bacterium in the human skin microbiome, as skin probiotics against methicillin-resistant *Staphylococcus aureus*. *PLoS ONE*, **8**, Available at: 10.1371/journal.pone.0055380. (Accessed 13/02/2014).

Sipos, R., Szekely, A. J., Palatinszky, M., Revesz, S., Marialigeti, K. and Nikolausz, M. (2007). Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology*, **60**, 341-350.

Smith, E. M., Monaghan, E. M., Huntley, S. J. and Green, L. E. (2011). Short communication: Preliminary investigation into the effect of freezing and a cryopreservant on the recovery of mastitis pathogens from ewe milk. *Journal of Dairy Science*, **94**, 4850-4855.

Supré, K., Haesebrouck, F., Zadoks, R. N., Vaneechoutte, M., Piepers, S. and De Vliegher, S. (2011). Some coagulase-negative *Staphylococcus* species affect udder health more than others. *Journal of Dairy Science*, **94**, 2329-2340.

Suzuki, M. and Giovannoni, S. (1996). Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*, **62**, 625-630.

Takeuchi, S., Nakajima, Y. and Hashimoto, K. (1983). Pathogenic synergism of *Fusobacterium necrophorum* and other bacteria in formation of liver-abcess in BALB/C mice. *Japanese Journal of Veterinary Science*, **45**, 775-781.

Tanaka, Y., Komaki, H., Yazawa, K., Mikami, Y., Nemoto, A., Tojyo, T., Kadowaki, K., Shigemori, H. and Kobayashi, J. (1997). Brasilinolide A, a new macrolide antibiotic produced by *Nocardia brasiliensis*: producing strain, isolation and biological activity. *The Journal of antibiotics*, **50**, 1036-1041.

Tanner, M. A., Goebel, B. M., Dojka, M. A. and Pace, N. R. (1998). Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Applied and Environmental Microbiology*, **64**, 3110-3113.

Taponen, S., Salmikivi, L., Simojoki, H., Koskinen, M. T. and Pyorala, S. (2009). Real-time polymerase chain reaction-based identification of bacteria in milk samples from bovine clinical mastitis with no growth in conventional culturing. *Journal of Dairy Science*, **92**, 2610-2617.

Tena, D., Martinez, N. M., Losa, C. and Solis, S. (2014). Skin and soft tissue infection caused by *Achromobacter xylosoxidans*: report of 14 cases. *Scandinavian Journal of Infectious Diseases*, **46**, 130-135.

Tilsala-Timisjärvi, A. and Alatossava, T. (2004). Rapid DNA preparation from milk and dairy process samples for the detection of bacteria by PCR. *International Journal of Food Microbiology*, **21**, 365-368.

Tola, S., Angioi, A., Rocchigiani, A. M., Idini, G., Manunta, D., Galleri, G. and Leori, G. (1997). Detection of *Mycoplasma agalactiae* in sheep milk samples by polymerase chain reaction. *Veterinary Microbiology*, **54**, 17-22.

Tomaso, H., Kattar, M., Eickhoff, M., Wernery, U., Al Dahouk, S., Straube, E., Neubauer, H. and Scholz, H. (2010). Comparison of commercial DNA preparation kits for the detection of Brucellae in tissue using quantitative real-time PCR. *BMC Infectious Diseases*, **10**, 1-5.

Trancassini, M., Iebba, V., Citera, N., Tuccio, V., Magni, A., Varesi, P., De Biase, R. V., Totino, V., Santangelo, F., Gagliardi, A. and Schippa, S. (2014). Outbreak of *Achromobacter xylosoxidans* in an italian Cystic fibrosis center: genome variability, biofilm production, antibiotic resistance, and motility in isolated strains. *Frontiers in Microbiology*, **5**, 8-16.

Tuke, P. W., Tettmar, K. I., Tamuri, A., Stoye, J. P. and Tedder, R. S. (2011). PCR Master Mixes Harbour Murine DNA Sequences. Caveat Emptor! *PLoS ONE*, **6**, Available at: 10.1371/journal.pone.0019953. (Accessed 02/05/14).

Turnbaugh, P. J., Ley, R. E., GHamady, M., Faser-Liggett, C. M., Knight, R. and Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, **449**, 804-810.

Twisk, J. W. R. (2003). Applied longitudinal data analysis for epidemiology. A practical guide, Volume 1, Cambridge University Press, UK.

van Schaik, G., Lotem, M. and Schukken, Y. H. (2002). Trends in somatic cell counts, bacterial counts, and antibiotic residue violations in New York State during 1999–2000. *Journal of Dairy Science*, **85**, 782-789.

Vela, A. I., Collins, M. D., Latre, M. V., Mateos, A., Moreno, M. A., Hutson, R., Dominguez, L. and Fernandez-Garayzabal, J. F. (2003). *Psychrobacter pulmonis* sp nov., isolated from the lungs of lambs. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 415-419.

Verdier-Metz, I., Michel, V., Delbès, C. and Christine Montel, M. (2009). Do milking practices influence the bacterial diversity of raw milk? *Food Microbiology*, **26**, 305-310.

Versalovic, J., Koeuth, T. and Lupski, J. R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Research*, **19**, 6823-31.

Watanabe, K., Nagao, N., Toda, T. and Kurosawa, N. (2008). Changes in bacterial communities accompanied by aggregation in a fed-batch composting reactor. *Current Microbiology*, **56**, 458-467.

Watkins, G. H., Burriel, A. R. and Jones, J. E. T. (1991). A field investigation of subclinical mastitis in sheep in southern England. *British Veterinary Journal*, **147**, 413-420.

Watson, D. and Buswell, J. (1984). Modern aspects of sheep mastitis. *British Veterinary Journal*, **140**, 529-34.

Watts, J. L. (1988). Etiological agents of bovine mastitis. *Veterinary Microbiology*, **16**, 41-66.

Witcomb, L. A., Green, L. E., Kaler, J., Ul-Hassan, A., Calvo-Bado, L. A., Medley, G. F., Grogono-Thomas, R. and Wellington, E. M. H. (2014). A longitudinal study of the role of *Dichelobacter nodosus* and *Fusobacterium necrophorum* load in initiation and severity of footrot in sheep. *Preventive Veterinary Medicine*, **115**, 48-55.

Yassin, A. F., Hupfer, H., Siering, C. and Schumann, P. (2011). Comparative chemotaxonomic and phylogenetic studies on the genus *Arcanobacterium Collins* et al. 1982 emend. Lehnen et al. 2006: proposal for *Trueperella* gen. nov and emended description of the genus *Arcanobacterium*. *International Journal of Systematic and Evolutionary Microbiology*, **61**, 1265-1274.

Yoon, J. H., Lee, K. C., Weiss, N., Kang, K. H. and Park, Y. H. (2003). *Jeotgalicoccus halotolerans* gen. nov., sp nov and *Jeotgalicoccus psychrophilus* sp nov., isolated from the traditional Korean fermented seafood jeotgal. *International Journal of Systematic and Evolutionary Microbiology*, **53**, 595-602.

Yu, K. F. and Pauls, K. P. (1992). Optimization of the PCR program for RAPD analysis. *Nucleic Acids Research*, **20**, 2606-2606.

Zadoks, R. N. and Watts, J. L. (2009). Species identification of coagulase-negative staphylococci: Genotyping is superior to phenotyping. *Veterinary Microbiology*, **134**, 20-28.

Zastempowska, E. and Lassa, H. (2012). Genotypic characterization and evaluation of an antibiotic resistance of *Trueperella pyogenes* (*Arcanobacterium pyogenes*) isolated from milk of dairy cows with clinical mastitis. *Veterinary Microbiology*, **161**, 153-158.

Appendix 1: SCC, teat lesion and microbiological culture data for sheep milk

Parity 1 sheep [38]

Sheep A20 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
279	L	2	396000	5.60	*Bacillus*, coliforms, *Staph*	S,S,S
567		3	94000	4.97	*Bacillus*, *Staph*	S,S
739		4	360000	5.56	*Bacillus*, G-ve, *Staph*	S,S,S
837		5	729000	5.86	coliforms, G-ve, *Staph*	S,S,S
1055		6	270000	5.43	*Bacillus*, *Staph*	S,S
1207		7	422000	5.63	*Bacillus*, *Staph*	S,S
1709		8	690000	5.84	*Bacillus*, *Staph*	S,S
280	R	1	216000	5.33	*Bacillus*, coliforms, *Staph*	S,S,S
568		2	76000	4.88	*Bacillus*, G-ve, *Staph*	S,S,S
740		3	404000	5.61	*Bacillus*, *Staph*	S,S
838		4	1130000	6.05	*Bacillus*, *Staph*	S,S
1056		5	450000	5.65	*Staph*	S
1208		6	1144000	6.06	*Bacillus*, *Staph*	S,S
1710		7	1072000	6.03	*Bacillus*, *Coryne*, *Staph*	S,S,S

Sheep A35 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
67	L	1	140000	5.15	*Bacillus*, maybe *Nocardia*, *Staph*	S,S,M
331		2	52000	4.72	*Bacillus*, maybe *Nocardia*	S,S
591		3	164000	5.21	-	NG
759		4	92000	4.96	*Bacillus*, *Alpha colonies*	S,S
823		5	58000	4.76	*Bacillus*, *Staph*	S,S
1073		6	208000	5.32	*Bacillus*, *Staph*	S,S
1185		7	254000	5.41	*Coryne*, *Staph*	S,S
332	R	1	76000	4.88	*Bacillus*, coliforms, *Coryne*, *Staph*	S,S,S,S
760		2	486000	5.69	*Alpha colonies*	S
824		3	98000	4.99	*Bacillus*, *Staph*	S,S
1074		4	336000	5.53	*Bacillus*, *Staph*	S,S
1186		5	336000	5.53	*Bacillus*, coliforms, fungal	M,M,S

[38] In all summary tables for sheep data in Appendix 1: *Staph* = *Staphylococcus* spp. *Coryne* = *Corynebacterium* spp., *Bacillus* = *Bacillus* spp., maybe *Nocardia* = maybe *Nocardia* spp., G-ve = Gram negative bacteria, fungal = fungal organism.
S = small growth of bacteria, M = medium growth of bacteria, H = high growth of bacteria, NG = no growth of bacteria.

Sheep A50 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
97	L	1	100000	5.00	*Bacillus*, G-ve, *Staph*	H,H,H
465		2	146000	5.16	*Bacillus*, G-ve, maybe *Nocardia*, *Staph*	S,M,S,H
681		3	48000	4.68	*Bacillus*, G-ve, *Staph*	S,S,H
777		4	76000	4.88	*Staph*	S
873		5	102000	5.01	*Bacillus*, *Staph*	S,S
1113		6	194000	5.29	*Bacillus*, coliforms, *Staph*, *Alpha colonies*	H,H,H,H
1235		7	90000	4.95	*Staph*	S
98	R	1	110000	5.04	*Bacillus*, *Staph*	H,H
224		2	160000	5.20	*Bacillus*, *Coryne*, *Staph*	S,S,S
466		3	948000	5.98	*Proteus*	H
874		6	138000	5.14	*Bacillus*, *Staph*	S,S
1114		7	1077000	6.03	*Staph*	S
1236		8	213000	5.33	*Bacillus*, *Coryne*	S,S

Parity 2 sheep

Sheep A15 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
27	L	1	236000	5.37	*Bacillus*, maybe *Nocardia*, *Staph*	S,S,M
299		2	92000	4.96	*Bacillus*	M
553		3	132000	5.12	*Bacillus*, G-ve, *Staph*	S,S,S
717		4	28000	4.45	*Staph*	S
841		5	268000	5.43	*Bacillus*	S
1047		6	420000	5.62	*Staph*	S
1223		7	390000	5.59	*Bacillus*, *Staph*	S,S
1291		8	1544000	6.19	*Bacillus*, *Coryne*	S,S
28	R	1	436000	5.64	*Bacillus*, *Staph*	S,S
300		2	62000	4.79	*Bacillus*, coliforms, *Staph*	S,S,S
554		3	94000	4.97	*Bacillus*, *Staph*	S,S
718		4	30000	4.48	*Staph*	S
842		5	192000	5.28	-	NG
1224		7	76000	4.88	*Bacillus*, *Staph*	S,S

Sheep A21 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
329	L	1	84000	4.92	*Bacillus*, coliforms, *Staph*	S,S,S
585		2	142000	5.15	*Bacillus*, maybe *Nocardia*, *Staph*	S,S,H
755		3	54000	4.73	*Coryne*, G-ve, *Staph*	S,S,S
819		4	148000	5.17	G-ve, *Staph*	S,S
1033		5	1642000	6.22	*Bacillus*	S
1159		6	116000	5.06	*Staph*	S
330	R	1	106000	5.03	*Bacillus*, coliforms, *Staph*	S,S,S
586		2	172000	5.24	*Bacillus*, *Coryne*, *Staph*	S,M,M
756		3	78000	4.89	*Bacillus*	S
820		4	156000	5.19	*Bacillus*, *Staph*	S,S
1034		5	1840000	6.27	*Bacillus*, *Staph*	S,S
1160		6	156000	5.19	*Bacillus*, *Staph*	S,S

Sheep A26 (No. of lambs reared = 3)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
225	L	1	60000	4.78	*Bacillus*, coliforms, *Staph*	S,S,S
457		2	93000	4.97	maybe *Nocardia*, *Staph*	S,S
669		3	26000	4.42	*Bacillus*, G-ve, *Staph*	S,S,S
775		4	54000	4.73	*Bacillus*, *Coryne*	S,S
875		5	172000	5.24	-	NG
1105		6	264000	5.42	*Coryne*, *Staph*	S,M
226	R	1	126000	5.10	*Bacillus*, *Coryne*, *Staph*	S,S,S
458		2	20000	4.30	*Bacillus*, *Staph*	M,H
670		3	70000	4.85	*Bacillus*, *Staph*	S,H
776		4	104000	5.02	*Bacillus*, maybe *Nocardia*, *Staph*	S,S,S
876		5	192000	5.28	G-ve	H
1106		6	160000	5.20	*Bacillus*	S
1238		7	450000	5.65	*Bacillus*, *Coryne*, *Staph*	S,S,S

Sheep A27 (No. of lambs reared = 3)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
455	L	1	206000	5.31	*Bacillus,* maybe *G-ve, Staph*	S,S,H
663		2	216000	5.33	maybe *Nocardia, Staph*	S,S
771		3	132000	5.12	*Bacillus, Staph*	S,S
1107		5	260000	5.42	*Bacillus, Staph, Alpha colonies*	S,S,S
1241		6	394000	5.60	*Bacillus, Staph*	S,S
456	R	1	276000	5.44	*Bacillus, Fungal, Staph*	S,M,M
664		2	108000	5.03	*Bacillus,* maybe *G-VE*	S,H
772		3	3000	3.48	*Bacillus, Coryne,* maybe G-ve, *Staph*	S,S,S,S
1108		5	566000	4.75	*Bacillus,* maybe *Nocardia, Staph*	S,S,S

Sheep A28 (No. of lambs reared = 3)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
589	L	1	176000	5.25	*Coryne, Staph*	S,S
761		2	126000	5.10	*Bacillus, Alpha colonies*	S,S
825		3	56000	4.75	*Staph*	S
1075		4	340000	5.53	-	NG
1183		5	152000	5.18	*Bacillus, Staph*	S,S
334	R	1	262000	5.42	*Bacillus,* maybe *Nocardia, Staph*	S,S,S
590		2	64000	4.81	*Bacillus, Coryne, Staph*	S,S,S
826		4	94000	4.97	G-ve, *Staph*	S,S
1076		5	291000	5.46	*Bacillus, Staph*	S,S
1184		6	110000	5.04	*Bacillus, Staph*	S,S

Sheep A39 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
75	L	1	372000	5.57	*Bacillus, Staph*	M,M
239		2	164000	5.22	*Coryne, Staph*	S,S
487		3	118000	5.07	*Bacillus, Staph*	S,S
697		4	40000	4.60	*Staph*	S
811		5	22000	4.34	*Staph*	S
1007		6	104000	5.02	-	NG
1139		7	106000	5.03	*Staph*	S
1273		8	94000	4.97	*Bacillus*	S
76	R	1	150000	5.18	*Bacillus, Staph*	S,S
240		2	368000	5.57	*Bacillus, Coryne, Staph*	S,S,S
488		3	150000	5.18	*Bacillus, Staph*	M,M
698		4	50000	4.70	*Bacillus,* G-ve, *Staph*	S,S,H
812		5	30000	4.48	*Staph*	S
1008		6	106000	5.03	*Bacillus, Staph*	S,S
1140		7	26000	4.42	*Staph*	S
1274		8	106000	5.03	*Bacillus, Staph*	S,S

Sheep A4 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
265	L	1	302000	5.48	*Bacillus*, coliforms, *Staph*	S,S,S
501		2	166000	5.22	*Bacillus*, fungal, *Staph*	S,S,S
703		3	116000	5.06	G-ve, *Staph*	S,S
795		4	220000	5.34	*Staph*	S
1127		5	184000	5.27	*Bacillus, Staph*	S,H
1271		6	890000	5.95	*Bacillus, Staph*	S,S
266	R	1	417000	5.62	*Bacillus, Staph*	S,S
502		2	374000	5.57	*Bacillus*, G-ve, *Staph*	S,S,S
704		3	144000	5.16	*Bacillus, Staph*	S,H
796		4	170000	5.23	*Bacillus*, G-ve, *Staph*	H,H,H

Sheep A41 (No. of lambs reared = 3)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
471	L	1	106000	5.03	*Bacillus*, maybe *Nocardia, Staph*	S,S,S
671		2	22000	4.34	*Bacillus, Staph*	S,H
779		3	62000	4.79	*Bacillus*	S
881		4	186000	5.27	*Bacillus*, G-ve, *Staph*	S,S,S
1109		5	86000	4.93	*Staph*	S
1245		6	504000	5.70	*Bacillus*	S
472	R	1	94000	4.97	*Bacillus, Staph*	S,M
672		2	14000	4.15	*Bacillus*, G-ve, *Staph, Alpha colonies*	M,S,H,H
780		3	58000	4.76	*Staph*	S
882		4	56000	4.75	*Bacillus*, G-ve, *Staph*	S,S,S
1246		6	222000	5.35	*Bacillus, Staph*	S,S

Sheep A44 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
559	L	1	68000	4.83	*Staph*	S
737		2	40000	4.60	*Staph*	S
863		3	48000	4.68	*Bacillus, Staph*	S,S
1067		4	254000	5.41	*Staph*	S
1201		5	94000	4.97	*Bacillus, Staph*	S,S
1287		6	284000	5.45	*Bacillus*, G-ve, *Staph*	S,H,H
282	R	1	274000	5.44	*Bacillus*, coliforms, *Coryne, Staph*	H,H,H,H
560		2	244000	5.39	*Bacillus*, G-ve	S,S
738		3	118000	5.07	*Staph*	S
864		4	318000	5.50	*Bacillus, Staph*	S,S
1068		5	212000	5.33	*Bacillus*	S
1202		6	112000	5.05	*Bacillus, Staph*	S,S
1288		7	634000	5.80	G-ve	S

Sheep A45 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
87	L	1	108000	5.03	*Bacillus, Staph*	S,S
319		2	450000	5.65	*Bacillus, Coryne, Staph*	M,S,M
581		3	140000	5.15	*Bacillus*	H
847		5	94000	4.97	*Bacillus, Staph*	S,S
1049		6	2217000	6.35	*Staph*	S
1205		7	176000	5.25	*Bacillus, Coliforms,* G-ve, *Staph*	S,S,S,S
582	R	1	138000	5.14	*Bacillus*	H
728		2	84000	4.92	*Bacillus,* G-ve, *Staph*	S,S,S
1050		4	684000	5.84	-	NG
1206		5	66000	4.82	*Bacillus, Coliforms,* fungal, *Staph*	S,S,S,S
1712		6	140000	5.15	*Bacillus*	S

Sheep A46 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
89	L	1	1146000	6.06	*Staph*	S
289		2	256000	5.41	*Bacillus,* fungal, *Staph*	M,S,S
575		3	80000	4.90	*Bacillus*	H
749		4	338000	5.53	*Staph*	H
853		5	78000	4.89	*Bacillus,* G-ve, *Staph*	S,S,S
1061		6	24000	4.38	*Bacillus, Staph*	S,S
1199		7	108000	5.03	*Bacillus,* fungal, *Staph*	S,S,M
1299	R	8	148000	5.17	*Staph*	H
90		1	1016000	6.01	*Bacillus,* maybe *Nocardia, Staph*	S,S,S
290		2	122000	5.09	*Bacillus,* maybe *Nocardia, Staph*	S,S,S
576		3	52000	4.72	*Bacillus*	H
750		4	84000	4.92	*Bacillus,* maybe *Nocardia, Staph*	S,S,H
854		5	58000	4.76	*Bacillus,* G-ve, *Staph*	S,S,S
1062		6	100000	5.00	*Bacillus, Coryne,* G-ve, *Staph*	M,M,M,M
1200		7	100000	5.00	*Bacillus, Staph*	S,S

Sheep A47 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
91	L	1	230000	5.36	*Bacillus, Coryne,* maybe *Nocardia, Staph*	S,S,S,S
269		2	58000	4.76	*Bacillus,* coliforms, *Coryne, Staph*	M,M,M,M
569		3	54000	4.73	*Bacillus,* G-ve, *Staph*	S,S,S
725		4	134000	5.13	*Bacillus, Staph*	S,S
859		5	174000	5.24	*Bacillus, Staph*	S,S,
1065		6	188000	5.27	*Staph*	S
1721		8	412000	5.62	*Staph*	S
92	R	1	220000	5.34	*Coryne, Staph*	S,S
270		2	160000	5.20	*Bacillus,* coliforms, *Staph*	M,M,M
570		3	50000	4.70	*Bacillus,* G-ve, *Staph*	S,S,S
726		4	58000	4.76	*Staph*	S
860		5	140000	5.15	*Staph*	S
1066		6	118000	5.07	*Bacillus, Staph*	S,S
1214		7	112000	5.05	*Bacillus, Staph*	S,S
1722		8	236000	5.37	*Staph*	S

Sheep A5 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
9	L	1	974	5.99	*Coryne, Staph, Alpha colonies*	M,M,M
327		2	492	5.69	*Bacillus,* maybe *Staph, Alpha colonies*	S,S,S
587		3	492	5.69	*Bacillus*	S
757		4	182	5.26	-	NG
821		5	286	5.46	-	NG
1031		6	370	5.57	*Bacillus,* maybe G-ve, *Staph*	S,S,S
1181		7	234	5.37	*Bacillus, Staph*	S,S
10	R	1	612	5.78	*Bacillus, Coryne, Staph*	M,M,M
328		2	303	5.48	*Bacillus, Coliforms*	S,S
588		3	548	5.74	*Coryne, Maybe Nocardia, Staph*	S,S,S
758		4	320	5.51	*Bacillus, Coryne,* maybe G-ve, *Alpha colonies*	M,H,S,S
822		5	300	5.48	*Bacillus, Staph*	S,S
1032		6	272	5.44	*Bacillus*	S
1182		7	360	5.56	*Bacillus*	S

Sheep A9 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
253	L	1	264000	5.422	*Bacillus, Coryne*	S,S
497		2	168000	5.225	*Bacillus*	S
807		4	46000	4.663	*Staph*	S
1009		5	164000	5.215	*Bacillus*	S
1137		6	120000	5.079	*Staph*	S
254	R	1	220000	5.342	-	NG
498		2	240000	5.380	*Bacillus, Staph*	S,S
694		3	52000	4.716	*Bacillus,* G-ve	S,H,H,
808		4	40000	4.602	*Staph*	S
1010		5	44000	4.643	*Bacillus, Staph*	S,S
1256		7	150000	5.176	*Staph*	S

Parity 3 sheep

Sheep A12 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
271	L	1	240000	5.38	*Bacillus*, coliforms, fungal, *Staph*	M,S,S,M
571		2	52000	4.72	*Coryne*	S
743		3	182000	5.26	-	NG
1063		5	160000	5.20	*Bacillus*	S
1217		6	96000	4.98	*Coryne, Staph*	S,S
1295		7	760000	5.88	*Staph*	S
272	R	1	243000	5.39	*Bacillus, Coryne,* fungal, *Staph*	M,S,S,M
572		2	76000	4.88	*Staph*	M
744		3	198000	5.30	-	NG
844		4	405000	5.61	*Bacillus, Staph*	S,S
1064		5	124000	5.09	*Bacillus, Staph*	S,S
1218		6	106000	5.03	*Coryne, Staph*	S,S
1296		7	342000	5.53	-	NG

Sheep A2 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
3	L	1	170	5.230	*Bacillus, Coryne*, maybe *Nocardia, Staph*	S,S,S,S
263		2	294	5.468	*Bacillus*, G-ve, *Staph*	S,S,S
483		3	164	5.215	*Staph*	S
793		5	242	5.384	*Bacillus, Staph*	S,S
1001		6	180	5.255	*Staph*	S
1131		7	302	5.480	-	NG
1259		8	1094	6.039	-	NG
4	R	1	910	5.959	*Bacillus, Staph*	S,H
264		2	2202	6.343	*Bacillus*, coliforms, *Staph*	M,M,M
484		3	738	5.868	*Bacillus, Staph*	S,M
696		4	873	5.941	*Bacillus*, G-ve, *Staph*	S,S,S
794		5	590	5.771	*Staph*	S
1002		6	872	5.941	-	-
1132		7	1268	6.103	*Bacillus, Staph*	S,H
1260		8	3069	6.487	*Coryne, Staph*	S,S

Sheep A22 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
325	L	1	240000	4.38	*Bacillus*, maybe *Nocardia*	M,S
579		2	218000	4.34	*Bacillus, Staph*	S,S
745		3	336000	4.53	*Bacillus, Staph*	S,S
861		4	134000	4.13	*Bacillus, Staph*	S,S
1071		5	124000	4.09	*Coryne, Staph*	S,M
1221		6	164000	4.22	*Staph*	S
1719		7	112000	4.05	*Staph*	S
326	R	1	160000	4.20	*Bacillus*, coliforms	S
580		2	202000	4.31	*Bacillus*	H
746		3	258000	4.41	*Staph*	S
862		4	148000	4.17	*Bacillus, Staph*	S,S

1072		5	76000	3.88	*Bacillus*, G-ve, *Staph*	M,M,M
1222		6	78000	3.89	*Bacillus, Staph*	S,S
1720		7	96000	3.98	-	NG

Sheep A25 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
47	L	1	172000	5.24	-	NG
287		2	126000	5.10	*Bacillus*, coliforms, *Staph*	S,S,S
565		3	20000	4.30	*Bacillus, Staph*	S,S
721		4	30000	4.48	*Coryne, Staph*	S,S
827		5	56000	4.75	*Bacillus*, G-ve, *Staph*	S,S,S
1045		6	32000	4.51	*Bacillus, Staph*	S,S
1189		7	40000	4.60	*Bacillus*	S
1293		8	240000	5.38	*Coryne, Staph*	S,S
288	R	1	11436000	7.06	*Bacillus*, coliforms, fungal	S,S,S
566		2	3854000	6.59	coliforms, *Staph*	H,S
722		3	2346000	6.37	coliforms	H
828		4	3582000	6.55	coliforms, *Staph*	M,S
1190		6	35432000	7.55	coliforms, *Staph*	M,S
1294		7	176000	5.25	*Bacillus, Coryne*, maybe *Nocardia*	S,S,S

Sheep A29 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
217	L	1	116000	5.06	*Bacillus, Proteus, Staph*	M,M,M
469		2	94000	4.97	-	NG
675		3	201000	5.30	*Bacillus, Staph*	S,H
773		4	66000	4.82	*Bacillus*, G-ve, *Staph, Alpha colonies*	M,M,M,M
877		5	146000	5.16	*Staph*	S
1101		6	94000	4.97	*Bacillus, Staph*	S,S
1231		7	684000	5.84	*Staph*	S
218	R	1	106000	5.03	*Bacillus, Staph*	M,M
470		2	42000	4.62	*Bacillus, Coryne*, maybe *Nocardia*	S,S,S
676		3	376000	5.58	*Staph*	S
774		4	52000	4.72	*Staph*	S
878		5	70000	4.85	-	NG
1102		6	144000	5.16	*Bacillus, Staph*	S,S
1232		7	748000	5.87	*Bacillus*	S

Sheep A3 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
261	L	1	160000	4.20	-	NG
803		2	30000	3.48	*Staph*	S
1019		3	92000	3.96	*Bacillus*, maybe *Nocardia*	S,S
1149		4	266000	4.43	*Staph*	S
1277		5	1032000	5.01	-	NG
544	R	1	88000	3.94	*Bacillus*	S
706		2	26000	3.43	*Staph*	S
804		3	24000	3.38	*Coryne*, G-VE, *Staph, Alpha colonies*	S,S,S,S
1020		4	72000	3.86	*Staph*	S

1150		5	292000	4.47	*Bacillus, Staph*	S,S
1278		6	1026000	5.01	*Bacillus,* maybe *Nocardia*	S,S

Parity 4 sheep

Sheep A16 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
317	L	1	172000	5.24	maybe *Nocardia*	S
557		2	14000	4.15	maybe *Nocardia*, *Staph*	S,S
715		3	26000	4.42	*Staph*	S
839		4	123000	5.09	*Bacillus*	S
1041		5	90000	4.95	*Bacillus*	S
1211		6	116000	5.06	coliforms, fungal, *Staph*	S,S,S
1297		7	446000	5.65	*Bacillus*, *Staph*	S,S
318	R	1	352000	5.55	*Bacillus*, *Coryne*, fungal, *Staph*	M,S,S,S
558		2	63000	4.80	*Staph*	S
716		3	58000	4.76	*Bacillus*, *Coryne*, G-ve, *Staph*	M,S,S,H
840		4	224000	5.35	*Bacillus*, *Staph*	H,H
1042		5	46000	4.66	*Bacillus*, *Staph*	S,S
1212		6	135000	5.13	*Bacillus*, *Staph*	S,S
1298		7	732000	5.87	*Bacillus*, G-ve, *Staph*	S,S,S

Sheep A24 (No. of lambs reared = 3)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
229	L	1	316000	5.50	*Bacillus*, maybe *Nocardia*, *Staph*	S,S,S
467		2	74000	4.87	*Bacillus*, *Staph*	S,S
677		3	192000	5.28	*Bacillus*, *Staph*	S,S
783		4	26000	4.42	*Bacillus*	S
885		5	124000	5.09	*Bacillus*, *Staph*	S,S
468	R	1	56000	4.75	*Bacillus*, G-ve, maybe *Nocardia*, *Staph*	S,S,S,H
678		2	206000	5.31	*Bacillus*, *Staph*	S,S
784		3	28000	4.45	*Bacillus*, *Coryne*, maybe *Nocardia*, *Staph*	S,S,S,S
886		4	78000	4.89	*Bacillus*, *Staph*	S,S
1244		6	256000	5.41	*Staph*	S

Sheep A40 (No. of lambs = 3)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
491	L	1	74000	4.87	*Bacillus*, *Staph*	S,S
691		2	54000	4.73	*Bacillus*, *Staph*	S,H
789		3	14000	4.15	*Bacillus*, maybe *Nocardia*	S,S
1003		4	104000	5.02	*Bacillus*	S
1133		5	432000	5.64	*Staph*	S
1265		6	294000	5.47	*Coryne*	S
250	R	1	120000	5.08	*Bacillus*, *Coryne*, maybe *Nocardia*, *Staph*	S,S,S,S
492		2	147000	5.17	*Bacillus*, G-ve, *Staph*	M,M,M
692		3	52000	4.72	*Staph*	S
790		4	18000	4.26	*Bacillus*, G-ve, *Staph*	H,H,H
1004		5	86000	4.93	-	NG

1134		6	270000	5.43	-	NG
1266		7	240000	5.38	*Coryne*	S

Sheep A43 (No. of lambs reared = 2)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
293	L	2	285000	5.46	*Bacillus, Staph*	S,S
547		3	207000	5.32	G-ve, *Staph*	S,S
735		4	82000	4.91	-	NG
829		5	98000	4.99	G-ve, *Staph, Alpha colonies*	S,S,S
1035		6	174000	5.24	*Staph*	S
1197		7	78000	4.89	*Bacillus,* fungal	S,S
1285		8	142000	5.15	G-ve, *Staph*	S,S
294	R	2	540000	5.73	*Bacillus,* maybe *Nocardia*	S,S
736		3	28000	4.45	-	NG
830		4	68000	4.83	*Bacillus,* G-ve, *Staph*	S,S,S
1036		5	192000	5.28	*Bacillus, Staph*	S,S
1198		6	46000	4.66	*Bacillus, Staph*	S,S
1286		7	180000	5.26	-	NG

Sheep A49 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
321	L	1	116	5.06	*Bacillus, Coryne*	S,S
555		2	60	4.78	*Bacillus,* coliforms, *Staph*	S,S,S
733		3	84	4.92	*Bacillus, Staph*	S,S
831		4	122	5.09	G-ve	S
1051		5	92	4.96	-	NG
1219		6	78	4.89	*Bacillus, Staph*	S,S
1713		7	204	5.31	*Bacillus, Coryne, Staph*	S,H,S
322	R	1	108	5.03	*Bacillus,* coliforms, *Staph*	S,S,S
832		4	148	5.17	*Bacillus, Staph*	S,S
1052		5	78	4.89	*Bacillus*	S
1220		6	68	4.83	*Staph*	S
1714		7	448	5.65	*Staph*	S

Parity 10 sheep

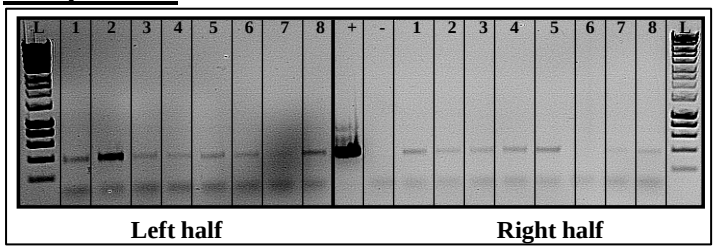
Sheep A6 (No. of lambs reared = unknown)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
11	L	1	628	5.80	*Bacillus, Staph*	S,S
699		2	26	4.42	*Bacillus, Staph*	S,H
787		3	32	4.51	*Bacillus, Staph*	S,S
1015		4	148	5.17	*Bacillus, Staph*	S,S
1135		5	76	4.88	*Staph*	S
1269		6	188	5.27	*Bacillus, Staph*	S,S
12	R	1	606	5.78	*Proteus*	H
700		2	104	5.02	*Bacillus, Staph*	S,S
788		3	48	4.68	*Bacillus, Staph*	S,S
1016		4	72	4.86	*Bacillus, Coryne, Staph*	S,S,S
1136		5	72	4.86	*Bacillus, Staph*	S,S
1270		6	147	5.17	*Bacillus, Coryne, Staph*	S,S,S

Sheep A23 (No. of lambs reared = 1)						
Sample No.	**Half**	**Week**	**SCC**	**Log SCC**	**Bacteria detected**	**Growth**
561	L	1	5358000	6.73	*Coryne, Staph*	S,S
751		2	3570000	6.55	*Bacillus*	S
855		3	3723000	6.57	*Bacillus*, G-ve, *Staph*	M,M,M
1043		4	3063000	6.49	*Bacillus*	S
1215		5	1956000	6.29	*Bacillus*, G-ve, *Staph*	S,S,S
1705		6	9084000	6.96	*Staph*	H
562	R	1	5256000	6.72	*Bacillus*, G-ve, *Staph*	S,S,S
752		2	12050000	7.08	-	NG
856		3	2372000	6.38	*Bacillus, Staph*	S,S
1044		4	1656000	6.22	*Bacillus, Staph*	S,S
1216		5	850000	5.93	-	-
1706		6	4578000	6.66	*Staph*	S

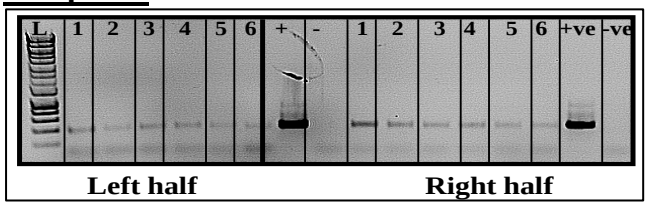
Appendix 2: Results of PCR amplification of sheep milk DNA

Parity 2 sheep

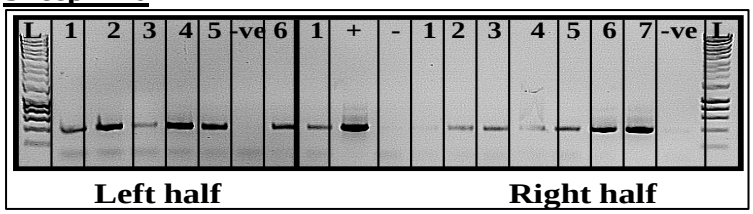
Sheep A15 [39]



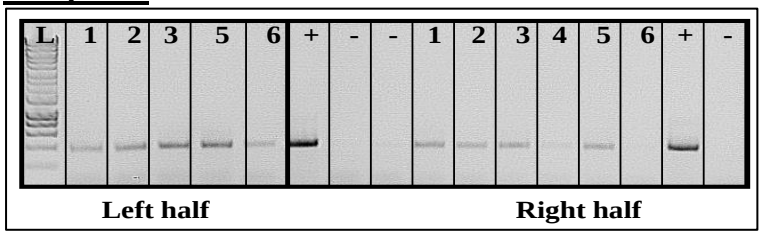
Sheep A21



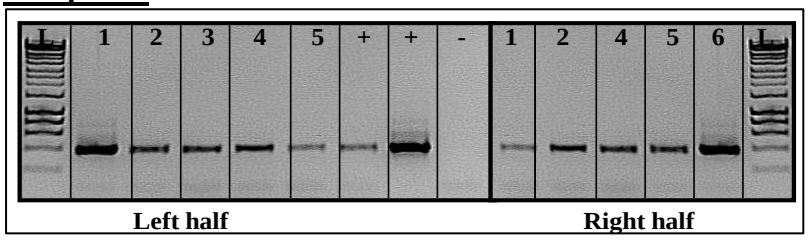
Sheep A26



Sheep A27

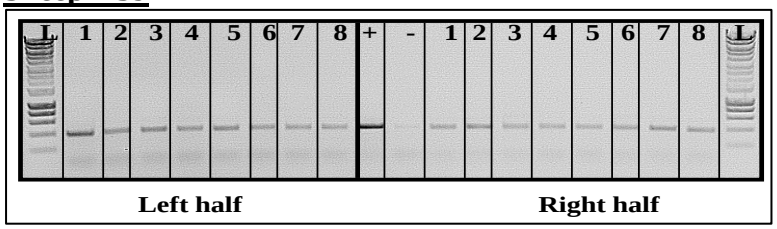


Sheep A28

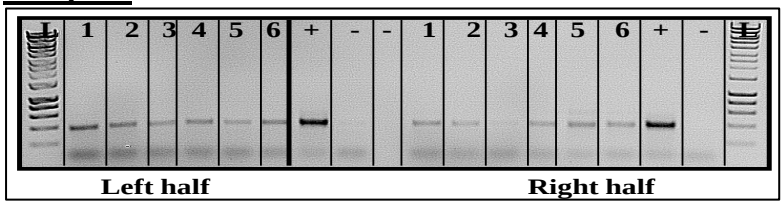


[39] In all Figures in Appendix 2, 'L' is Hyperladder 1kb (Bioline, UK); numbers 1-8 are weeks 1-8 of the study sampling period and the mammary gland half the weeks relate to is shown in each Figure. The '+' and '-' are the DNA extraction positive and negative controls respectively.

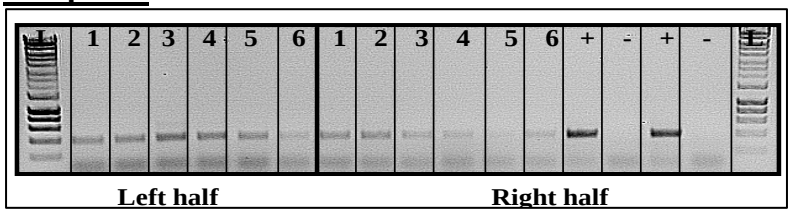
Sheep A39



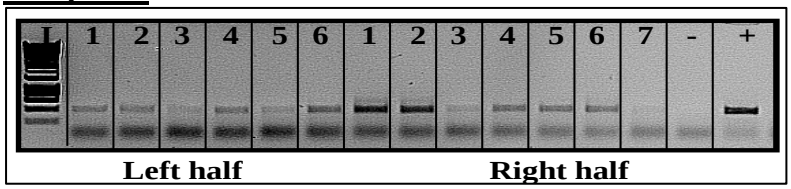
Sheep A4



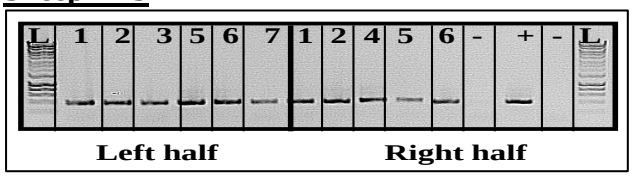
Sheep A41



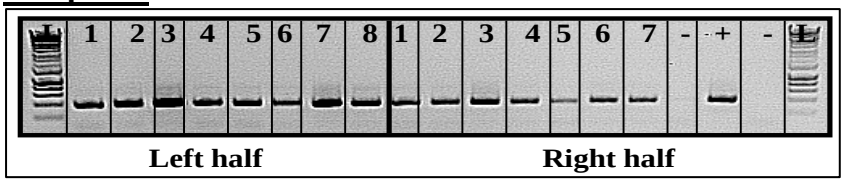
Sheep A44



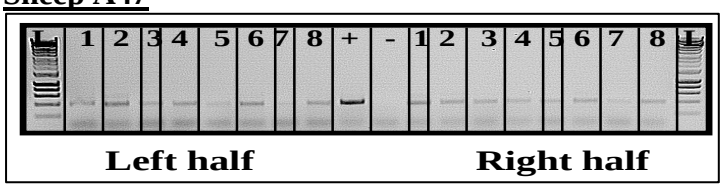
Sheep A45



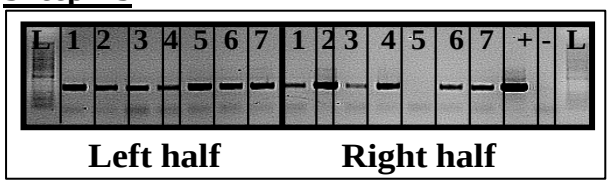
Sheep A46



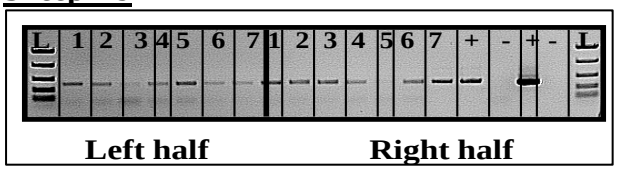
Sheep A47



Sheep A5

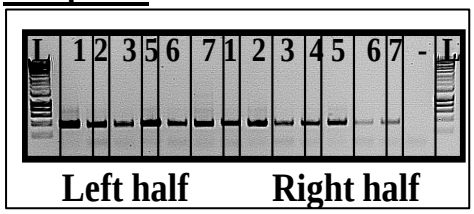


Sheep A9

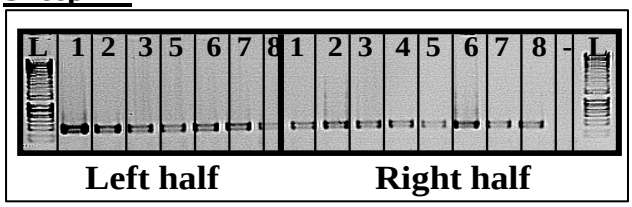


Parity 3 sheep

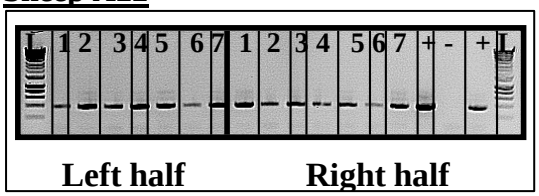
Sheep A12



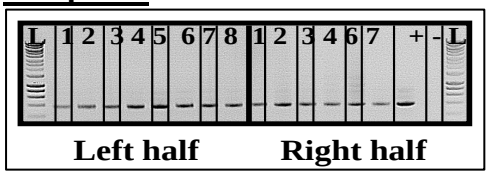
Sheep A2



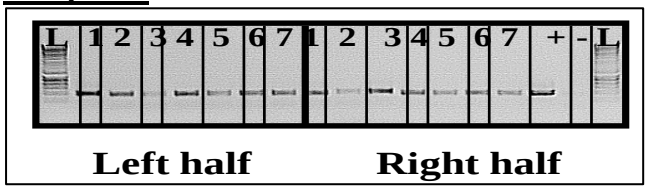
Sheep A22



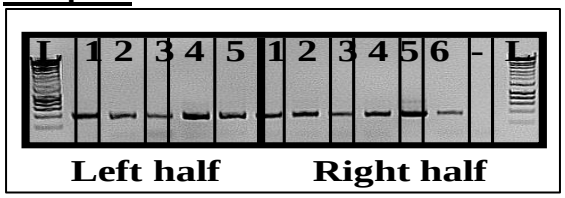
Sheep A25



Sheep A29

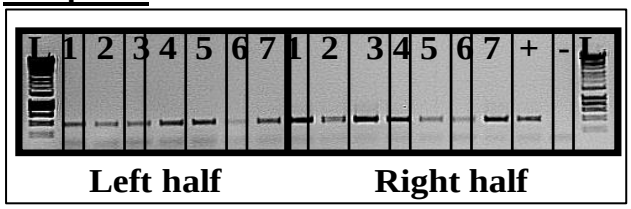


Sheep A3

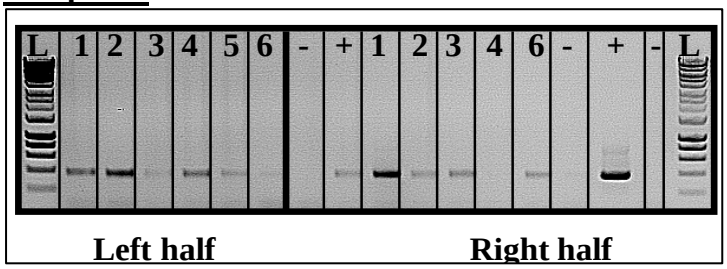


Parity 4 sheep

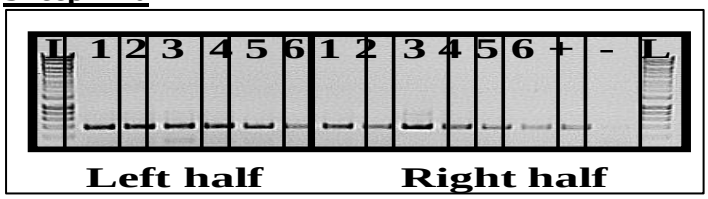
Sheep A16



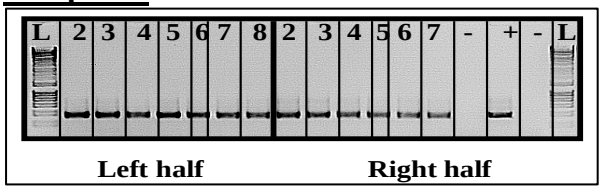
Sheep A24



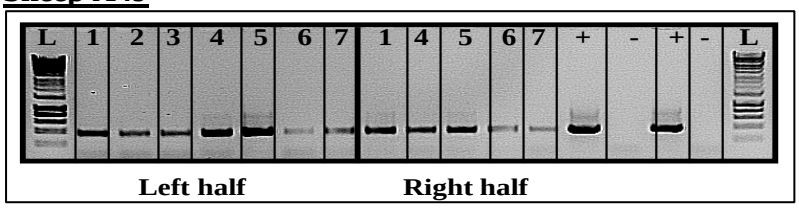
Sheep A40



Sheep A43

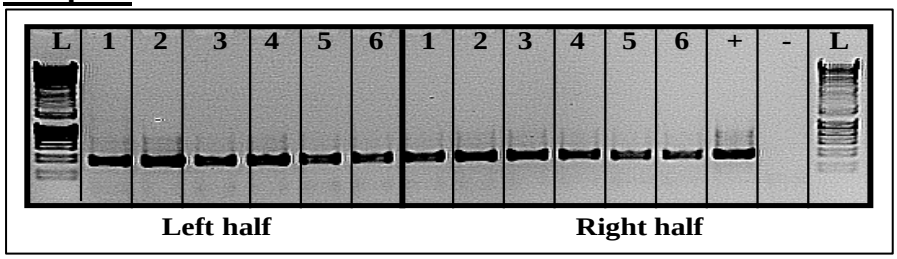


Sheep A49

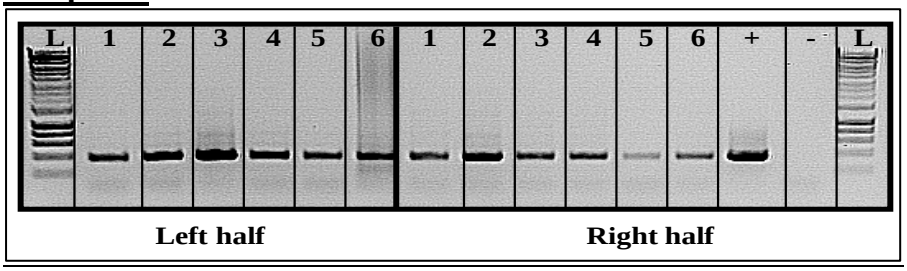


Parity 10 sheep

Sheep A6



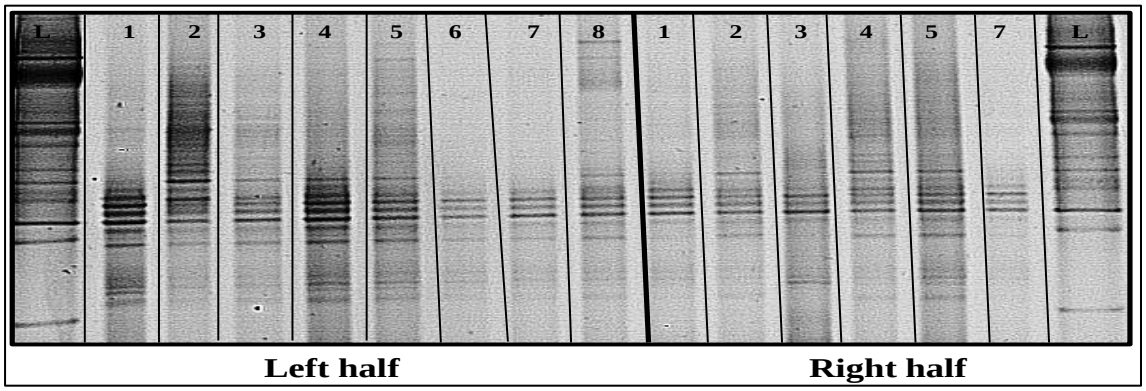
Sheep A23



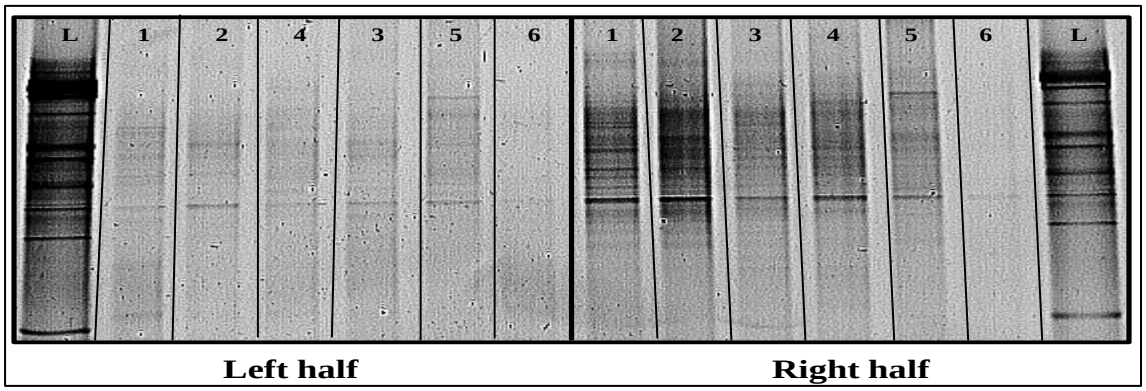
Appendix 3: DGGE images for sheep milk samples

DGGE results for parity 2 sheep

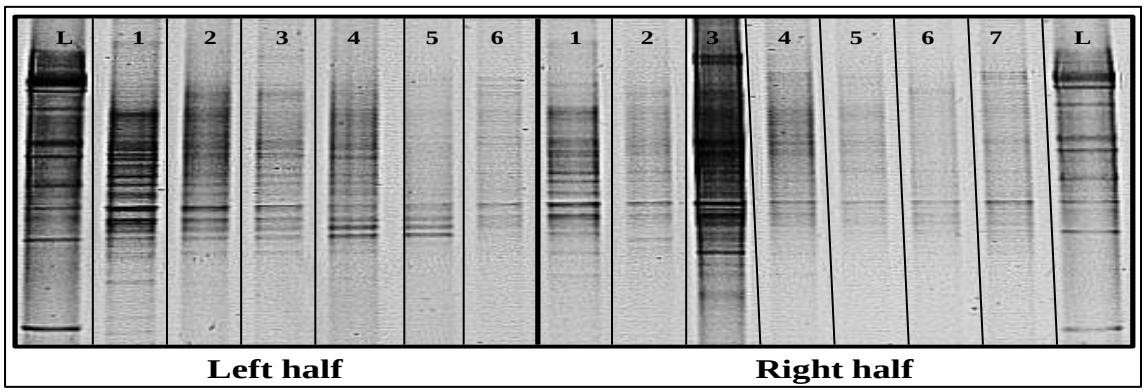
Sheep A15: [40]



Sheep A21:

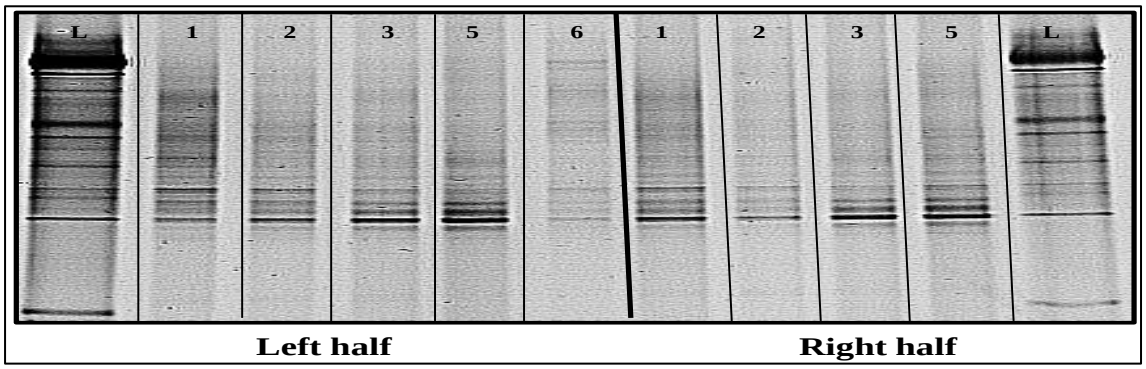


Sheep A26:

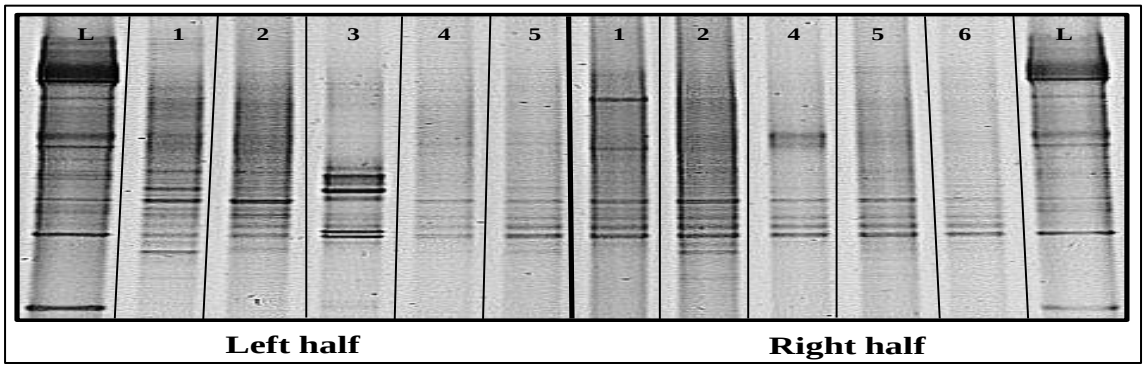


[40] For all DGGE images, 'L' corresponds to the custom DGGE reference ladder; numbers 1-8 are milk samples from weeks 1-8, with the mammary gland half they originate from specified in each Figure.

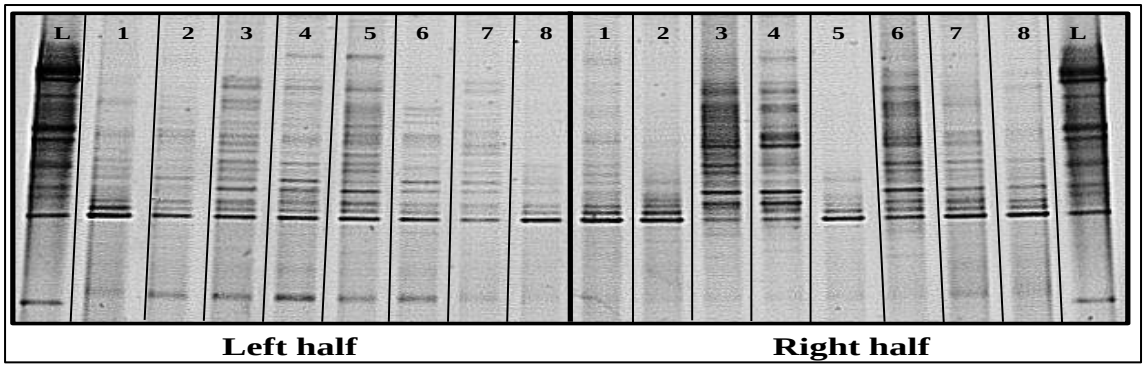
Sheep A27:



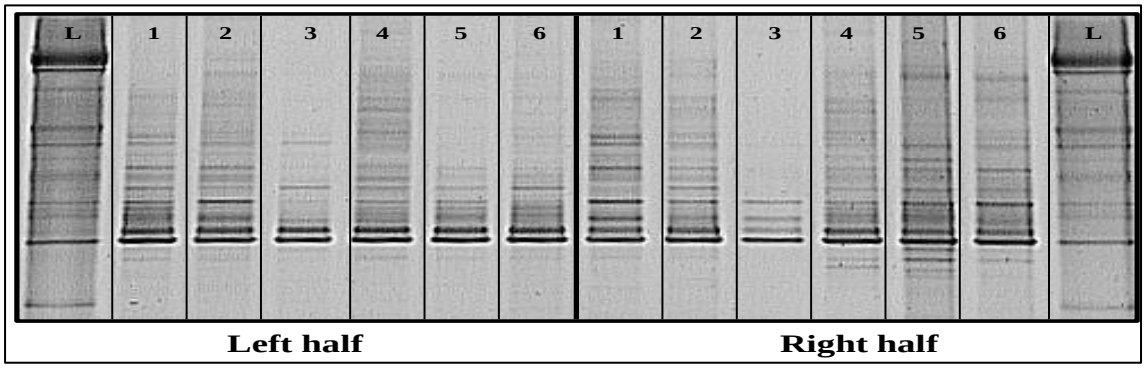
Sheep A28:



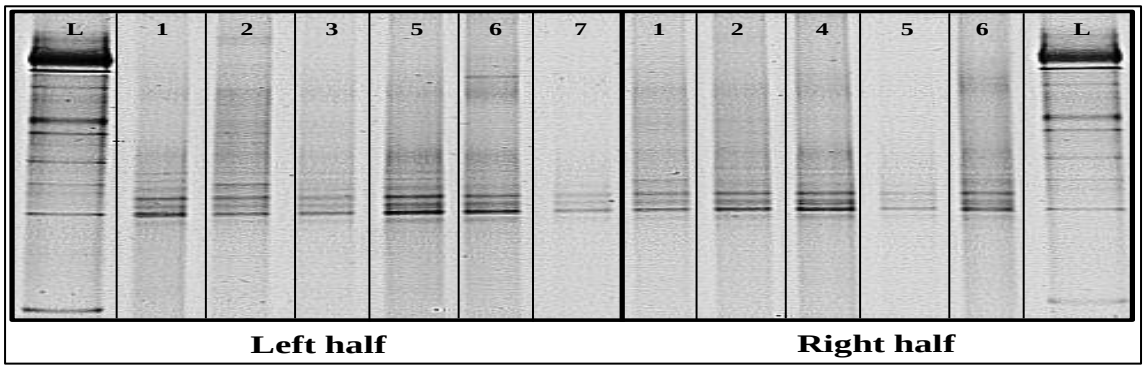
Sheep A39:



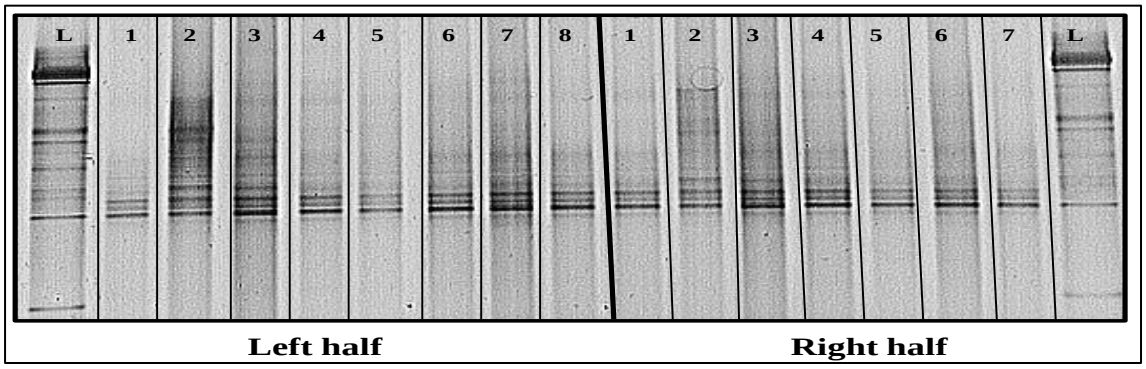
Sheep A4:



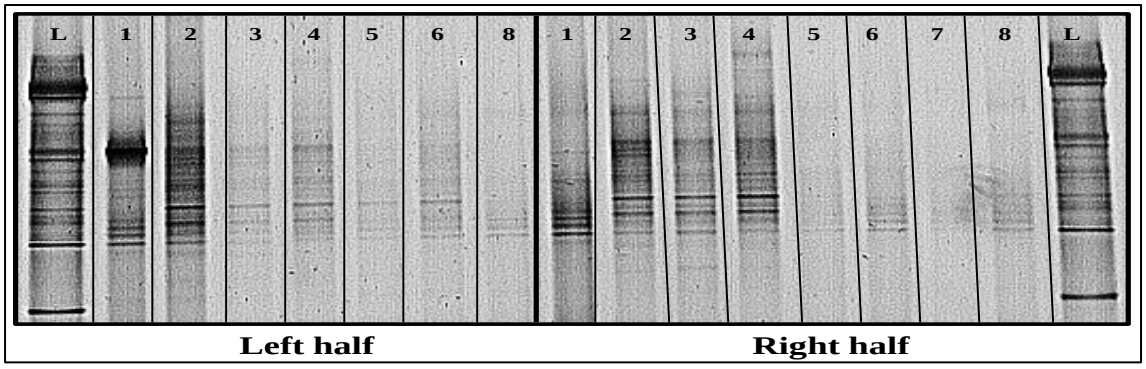
Sheep A45:



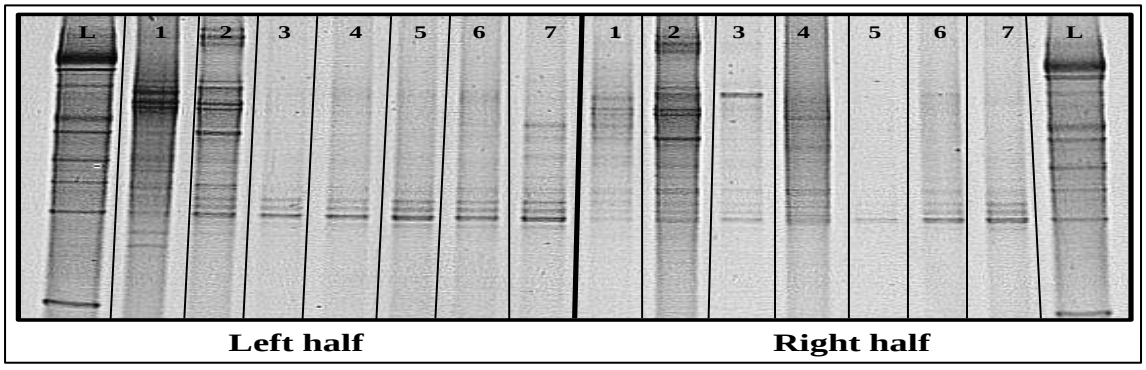
Sheep A46:



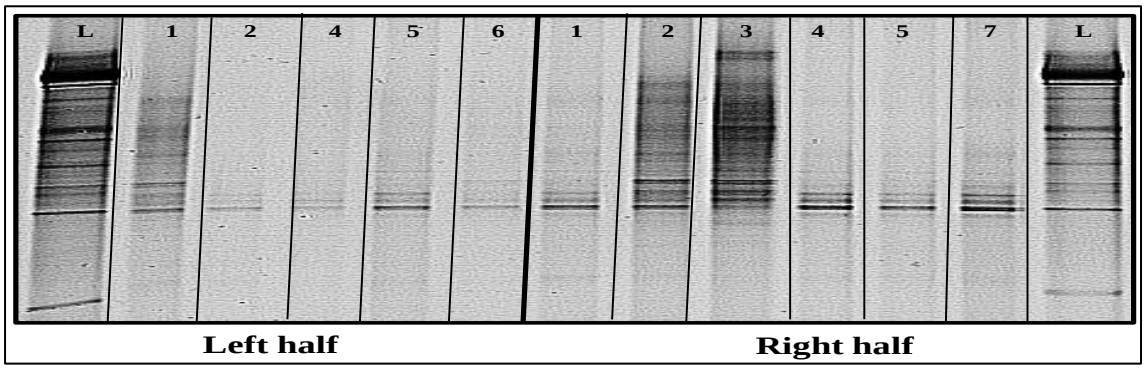
Sheep A47:



Sheep A5:

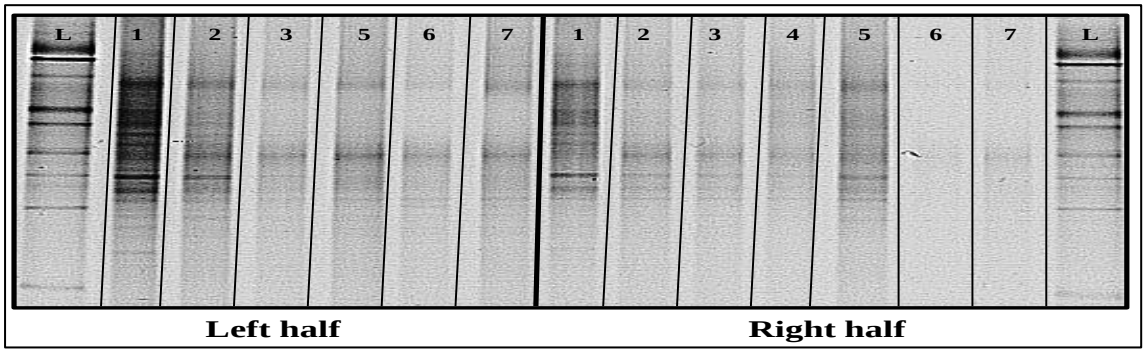


Sheep A9:

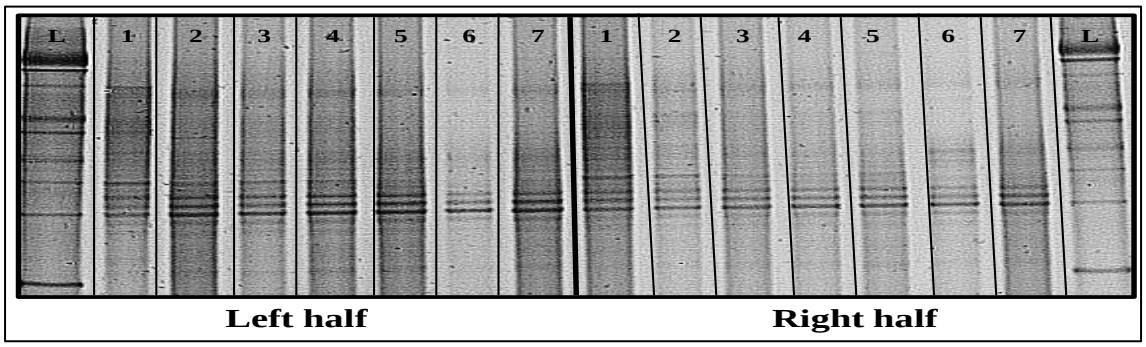


DGGE results for parity 3 sheep

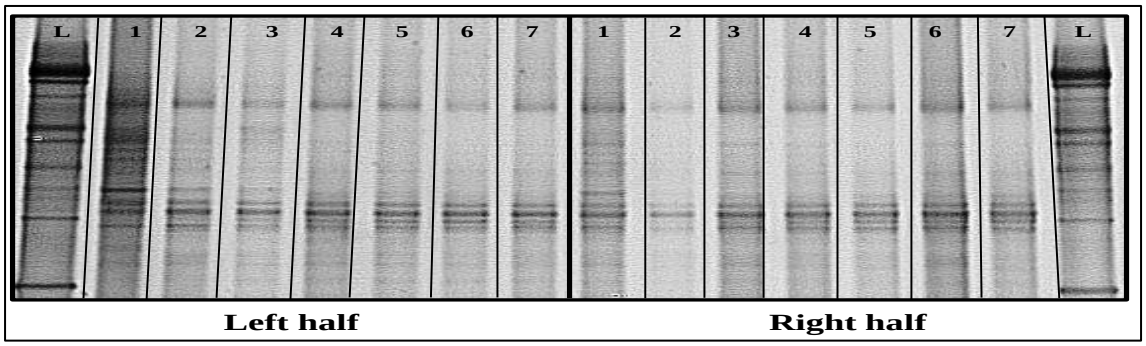
Sheep A12:



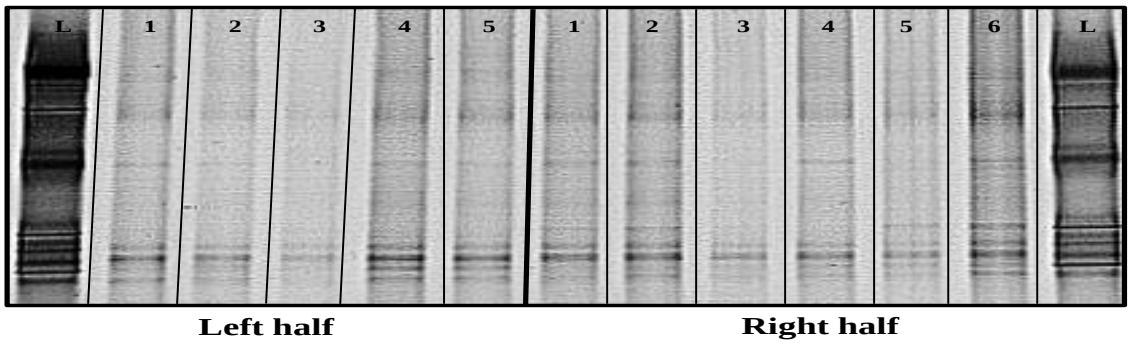
Sheep A22:



Sheep A29:

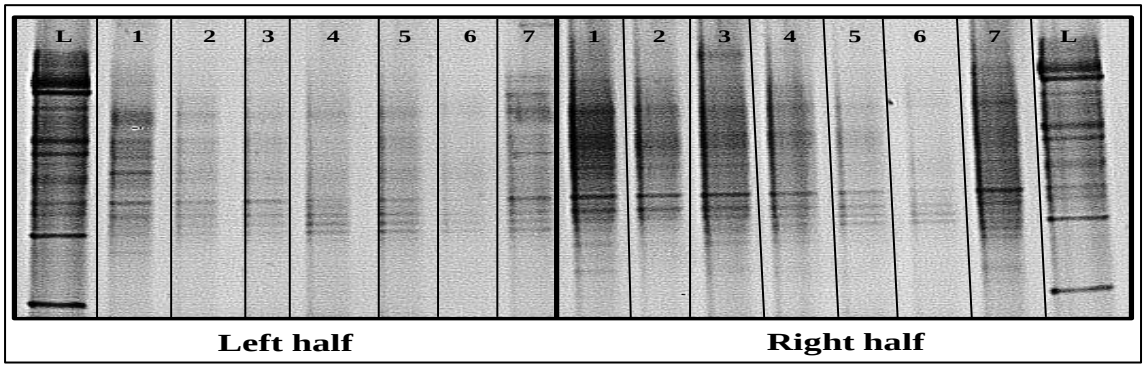


Sheep A3:

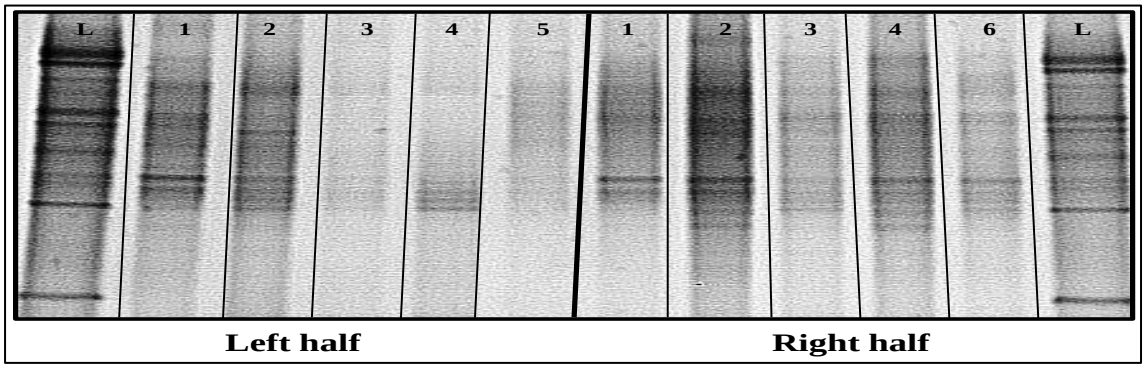


DGGE results for parity 4 sheep

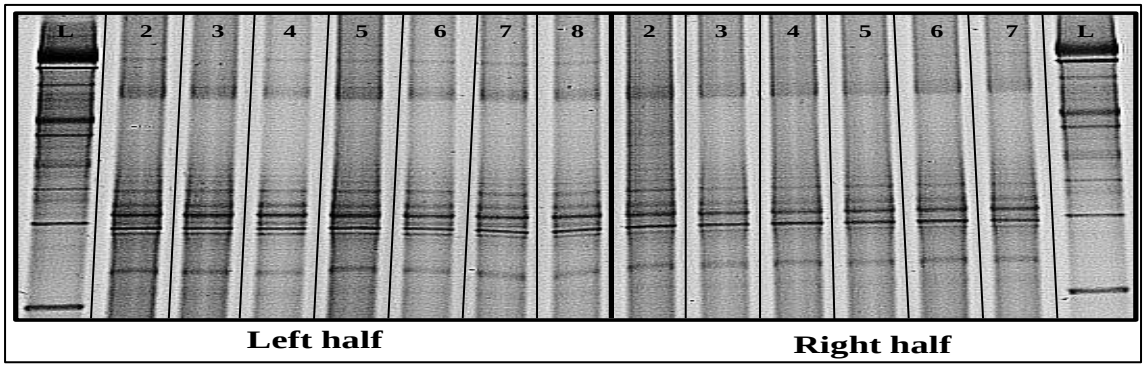
Sheep A16:



Sheep A24:



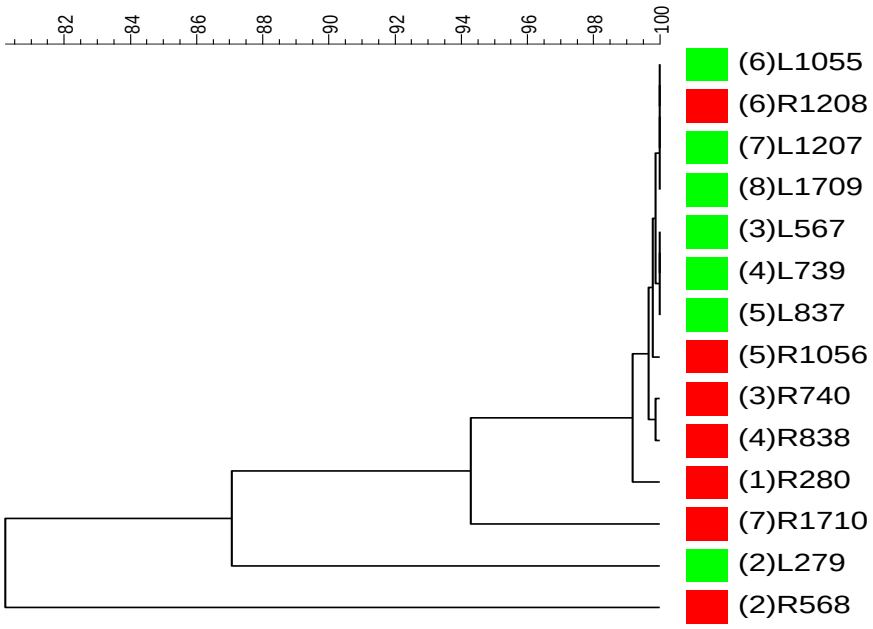
Sheep A43:



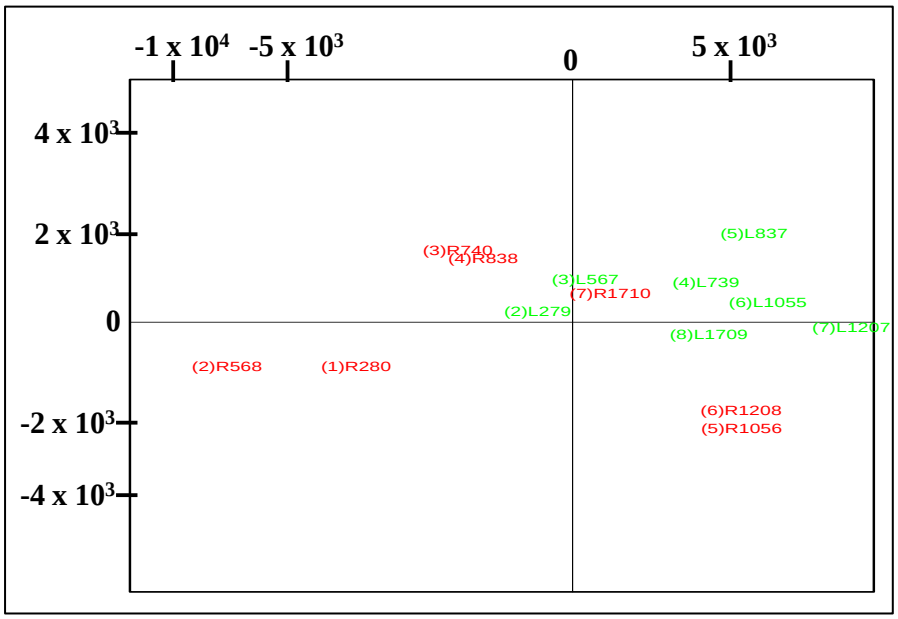
Appendix 4: Dendrograms and PCA plots for milk samples grouped by sheep

Parity 1 sheep

Dendrogram of sheep A20 milk samples [41]

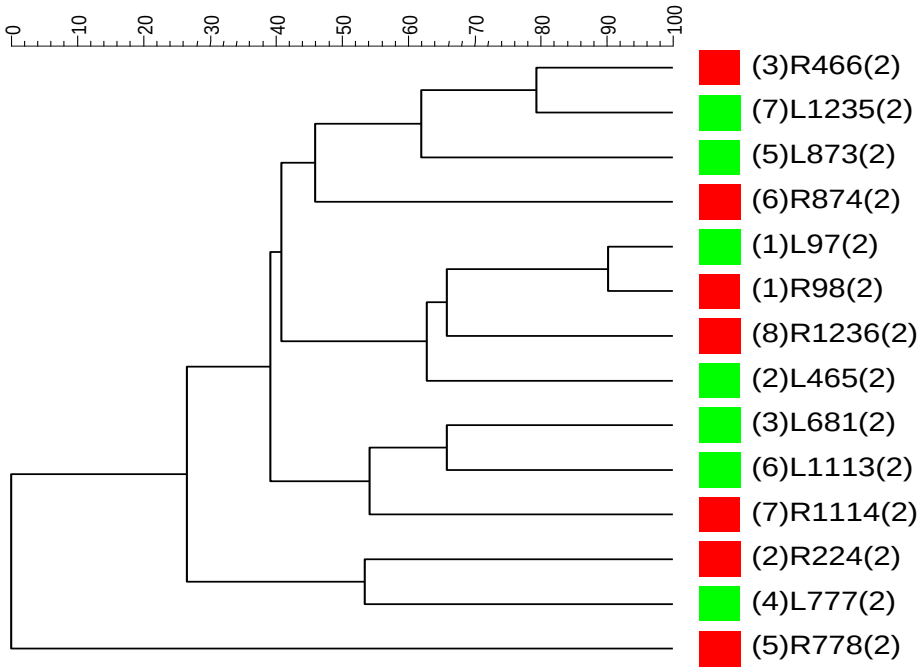


PCA plot of sheep A20 milk samples

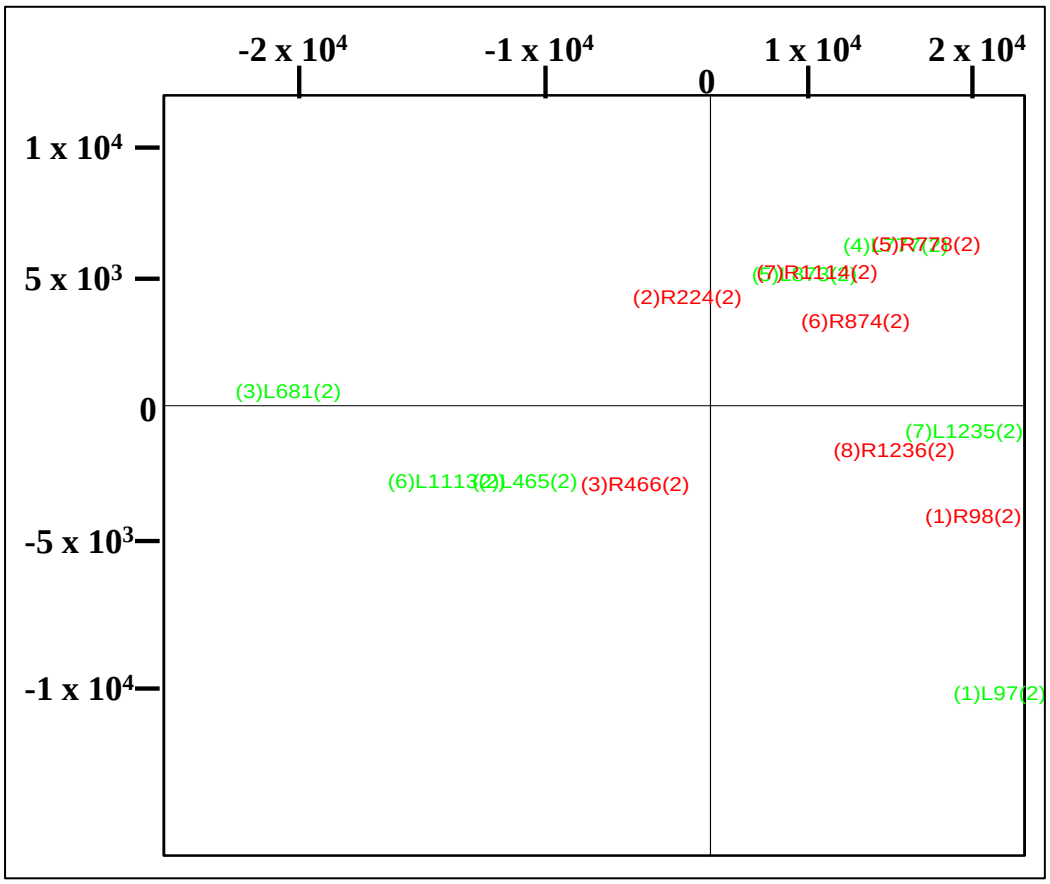


[41] All dendrogram axis are the percentage similarity. The lowest percentage is the greatest similarity between the 2 most dissimilar samples in the dendrogram. The axis on the PCA plots are the entry coordinates which are the coordinates of the entries in the first 2 components of the PCA.

Dendrogram of sheep A50 milk samples

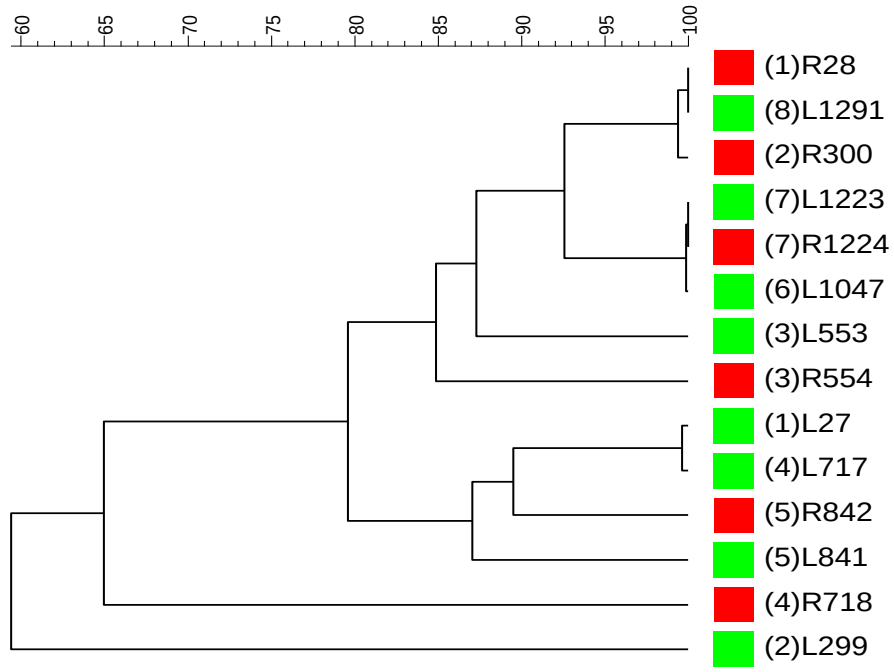


PCA plot of sheep A50 milk samples

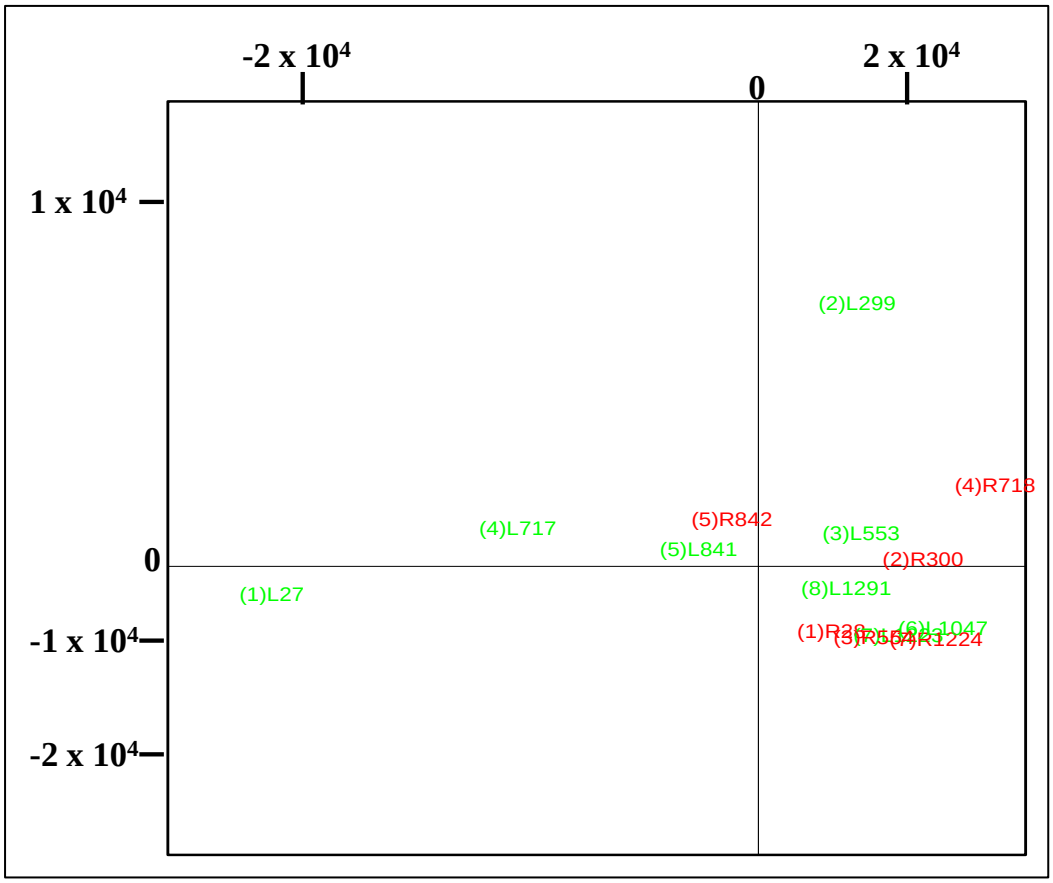


Parity 2 sheep

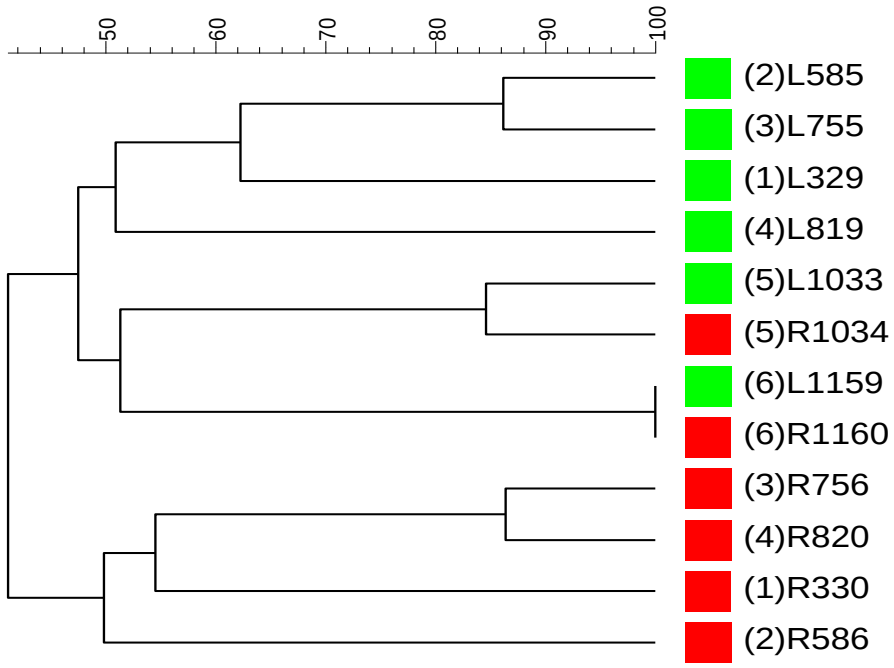
Dendrogram of sheep A15 milk samples



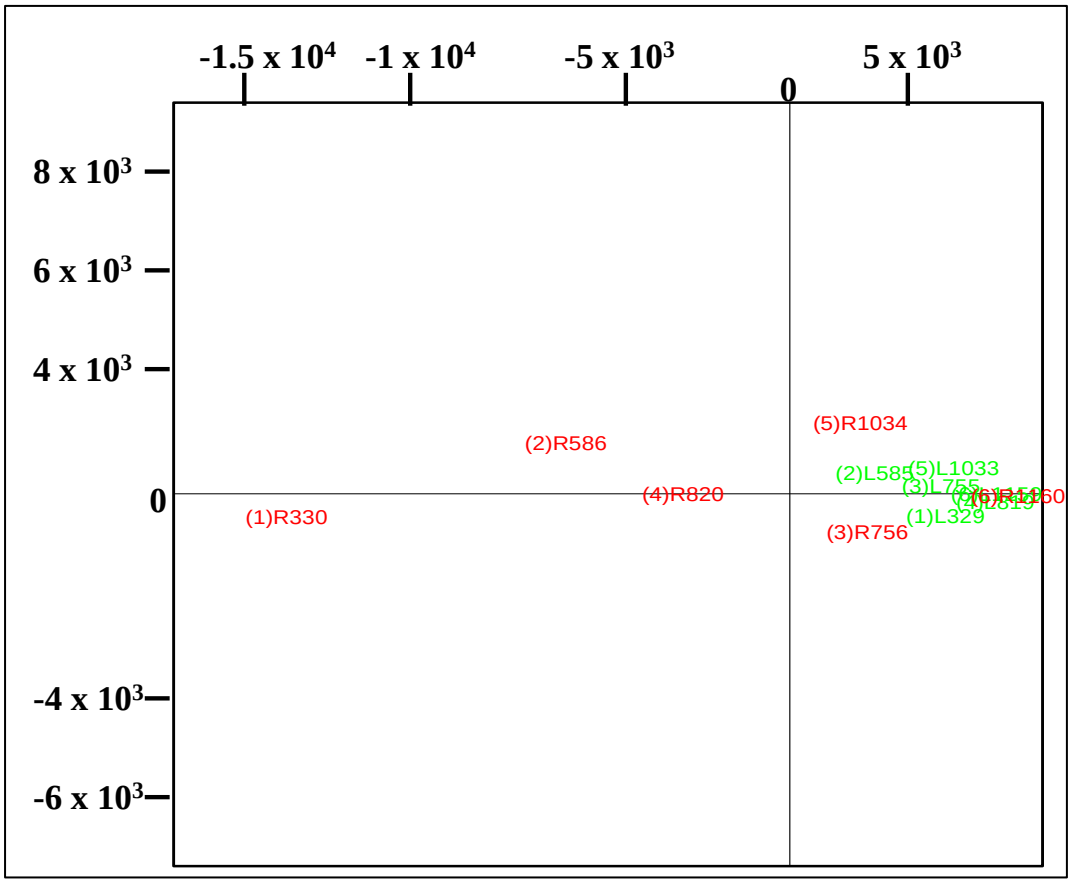
PCA plot of sheep A15 milk samples



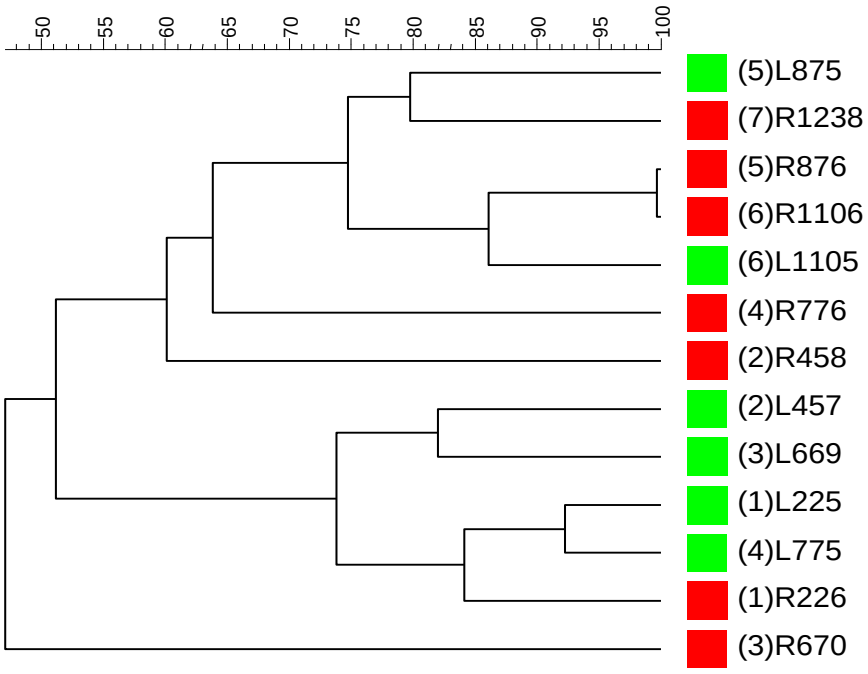
Dendrogram of sheep A21 milk samples



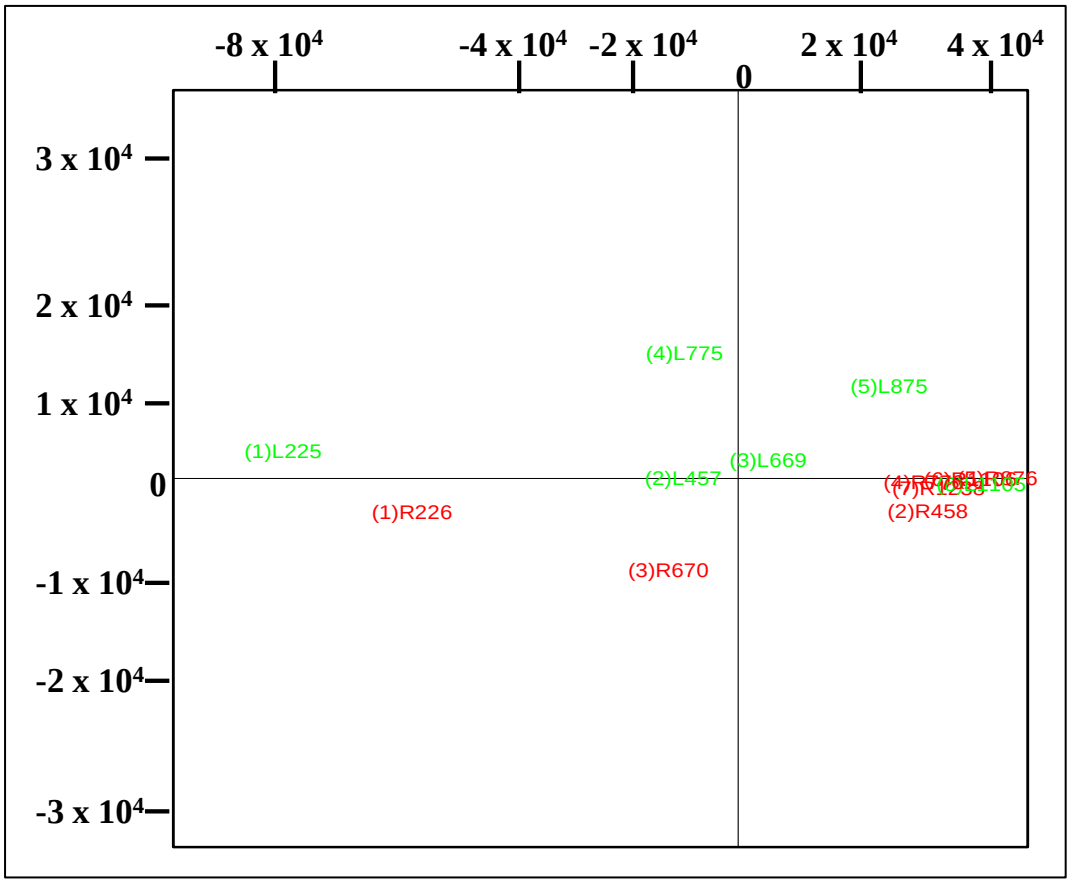
PCA plot of sheep A21 milk samples



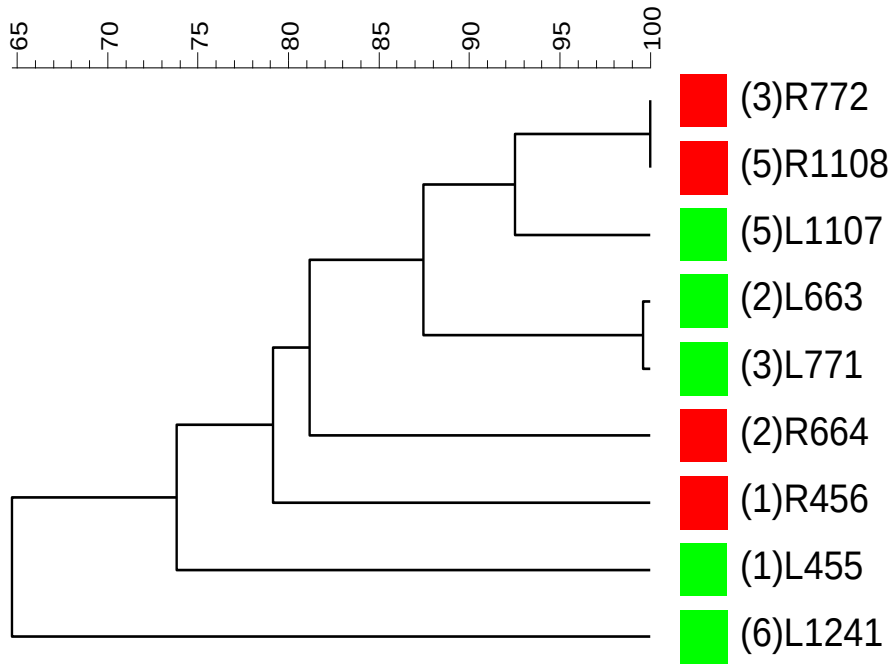
Dendrogram of sheep A26 milk samples



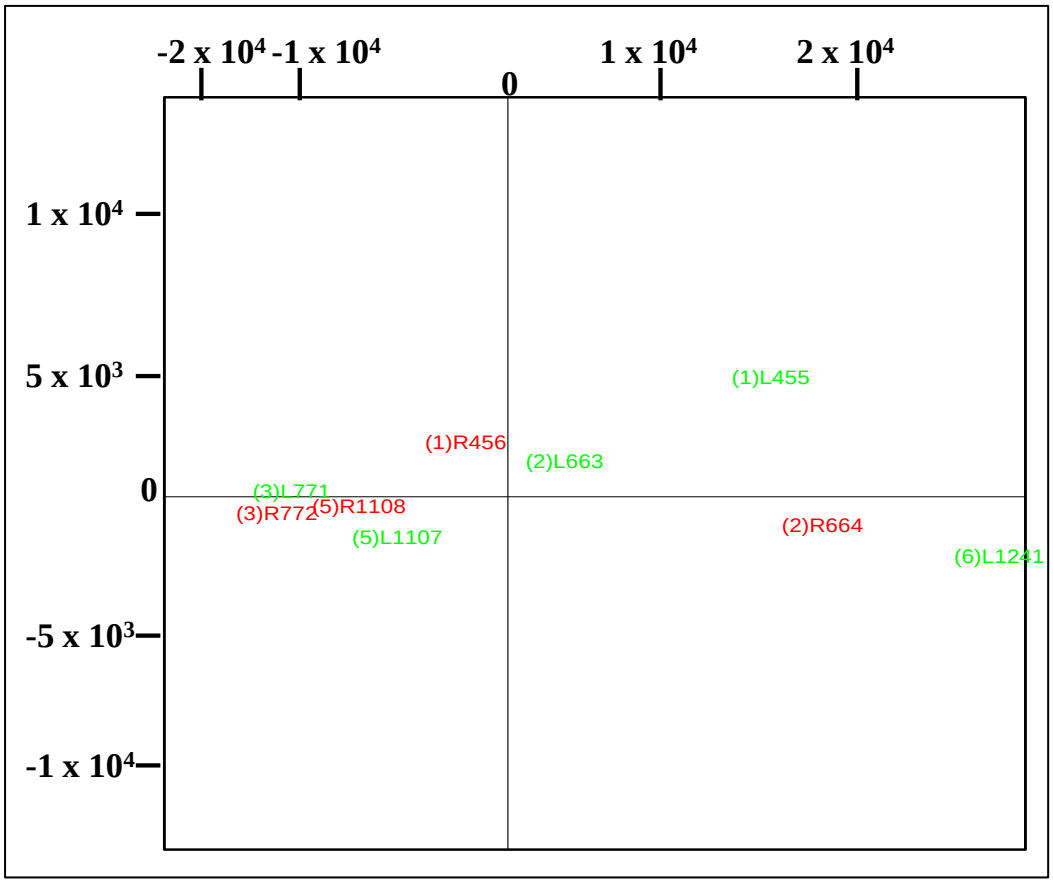
PCA plot of sheep A26 milk samples



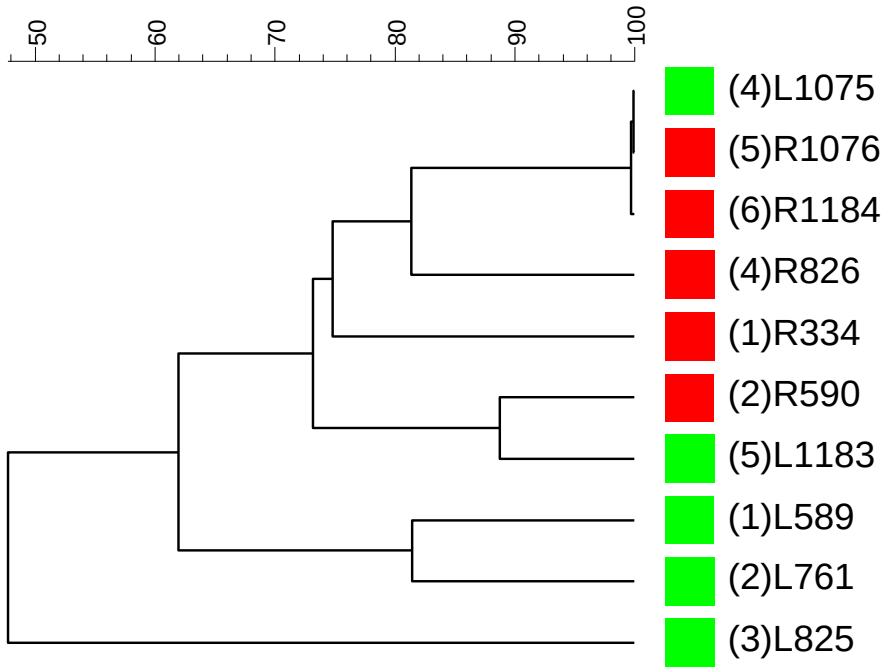
Dendrogram of sheep A27 milk samples



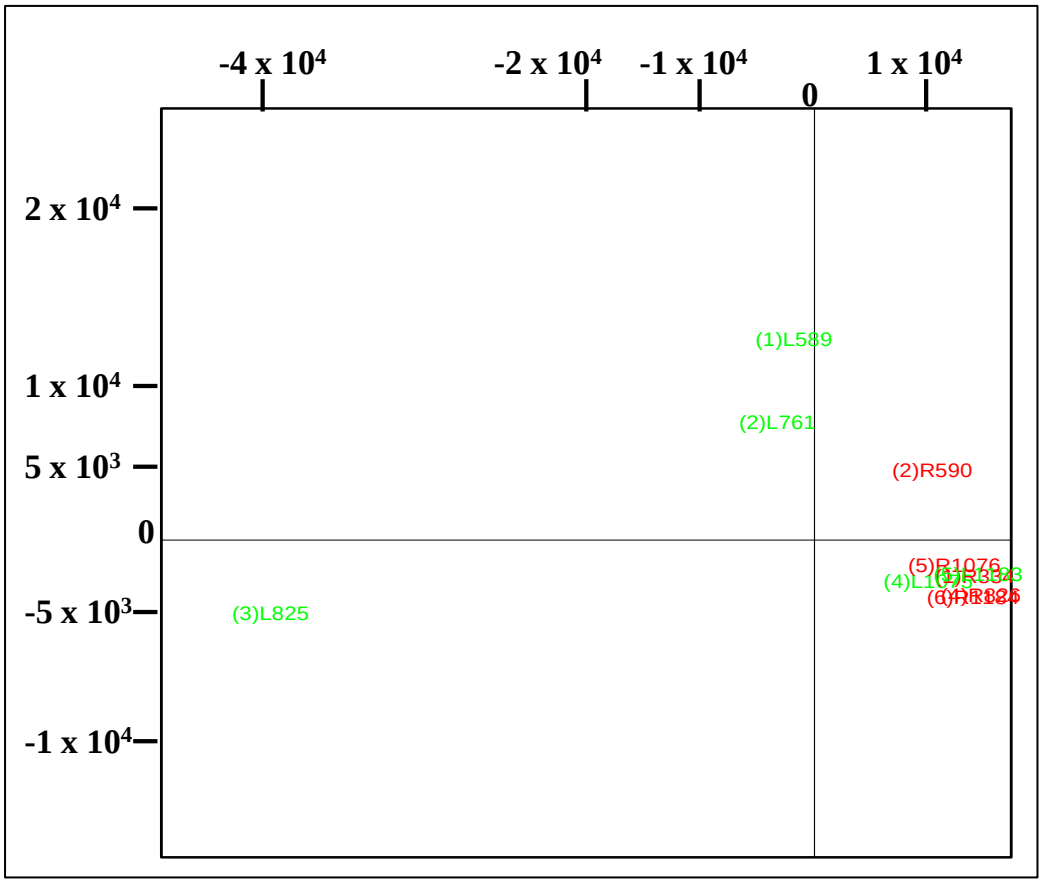
PCA plot of sheep A27 milk samples



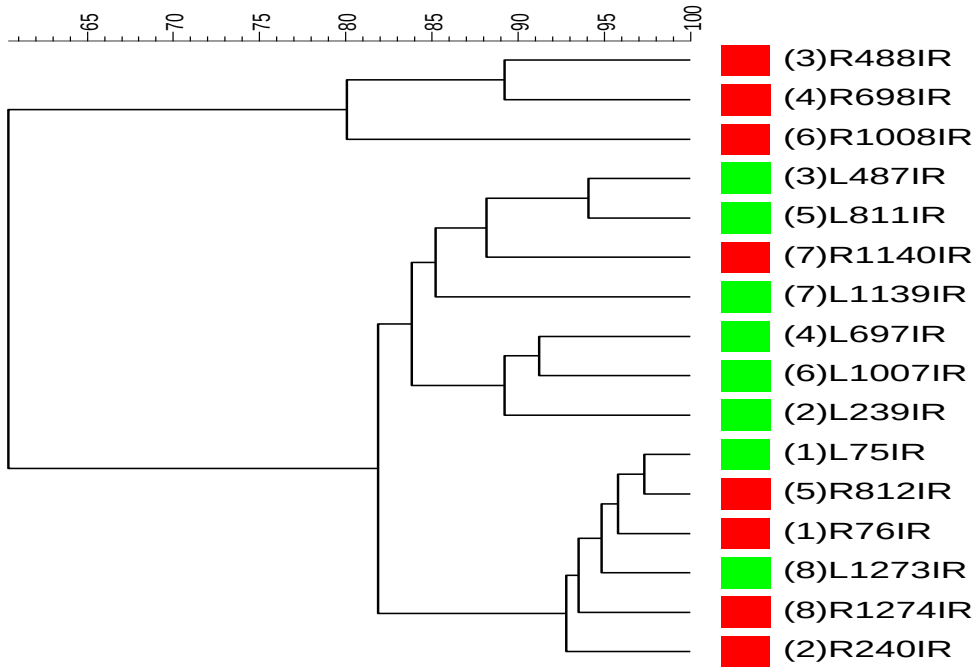
Dendrogram of sheep A28 milk samples



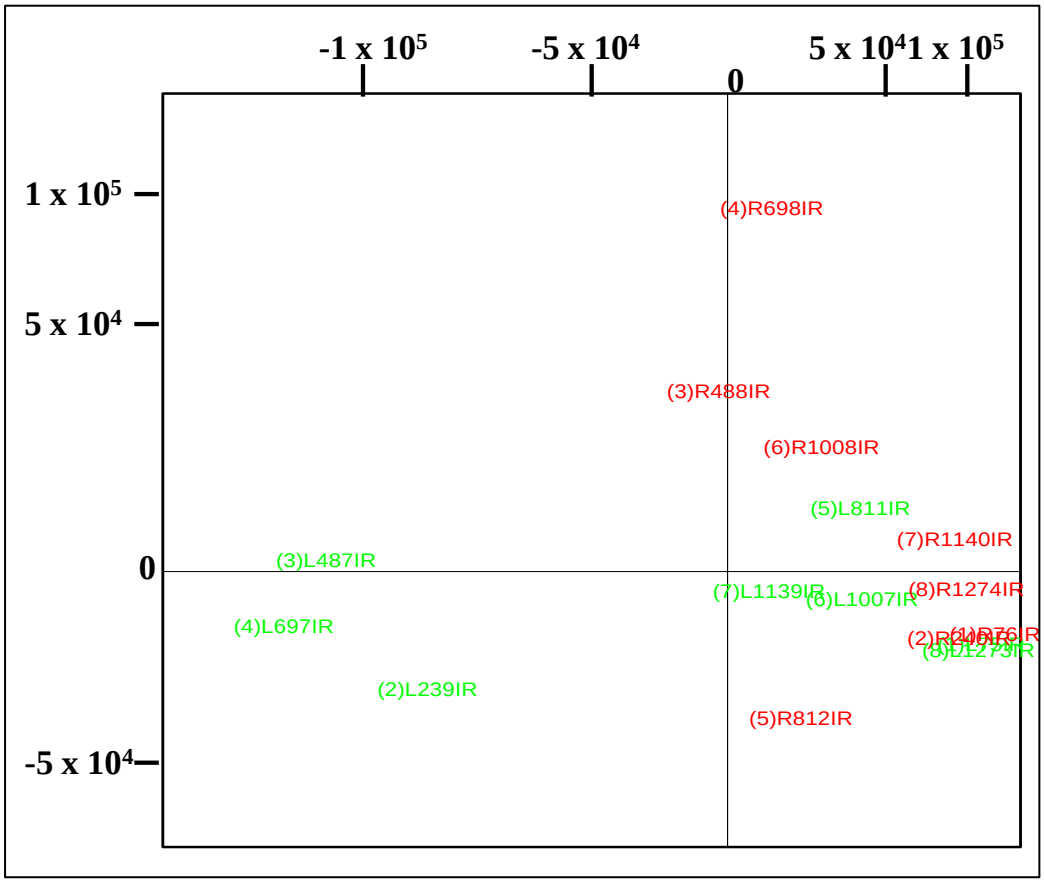
PCA plot of sheep A28 milk samples



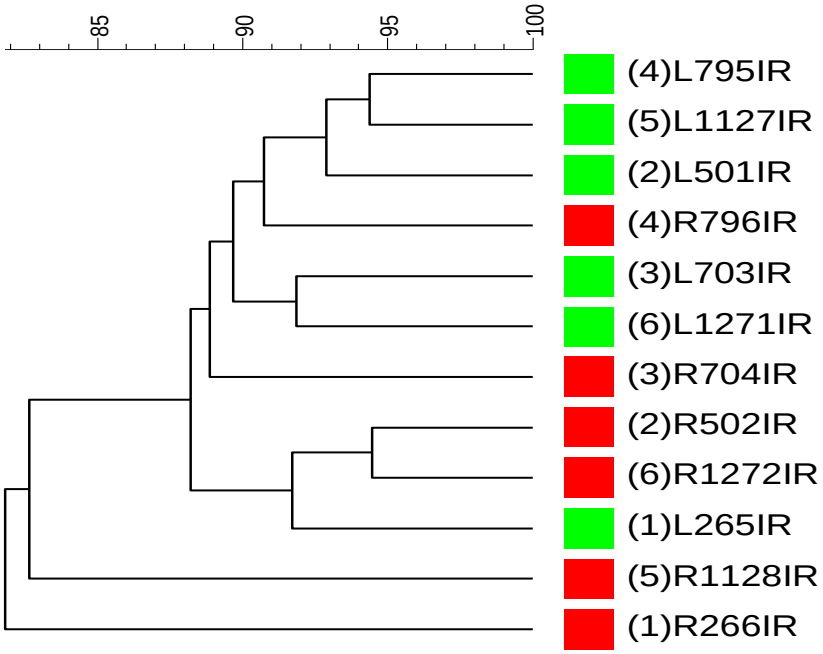
Dendrogram of sheep A39 milk samples



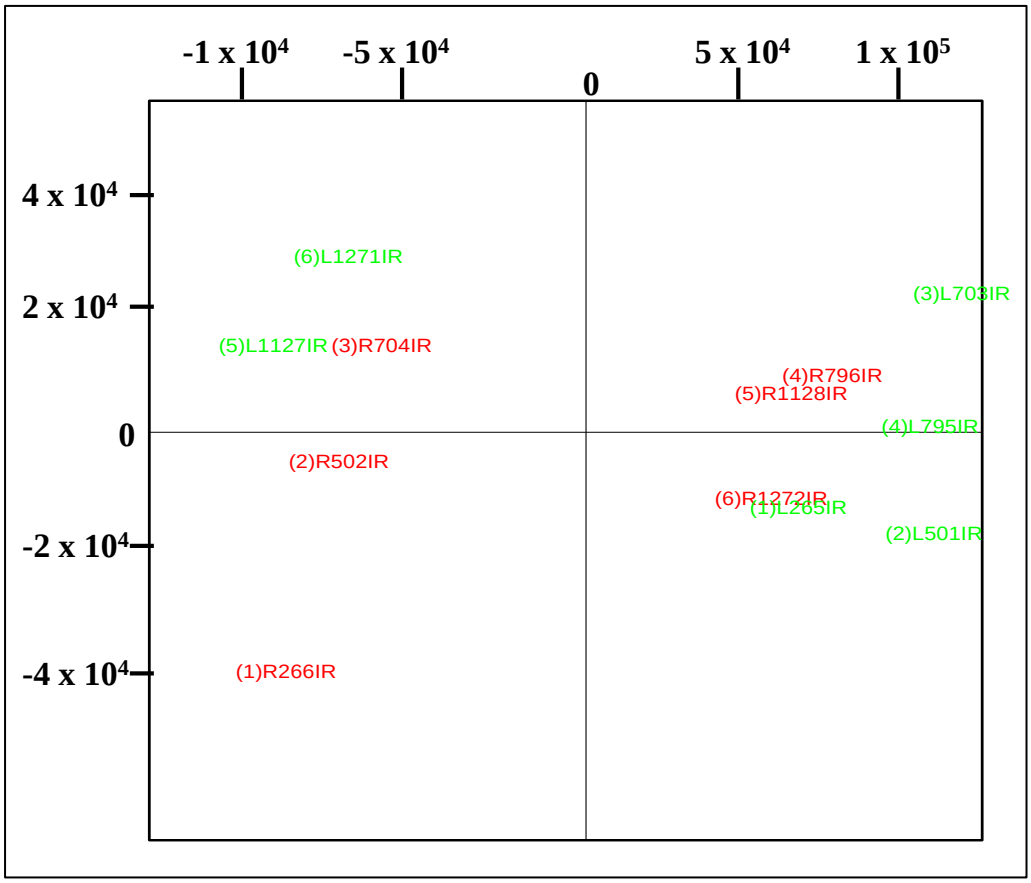
PCA plot of sheep A39 milk samples



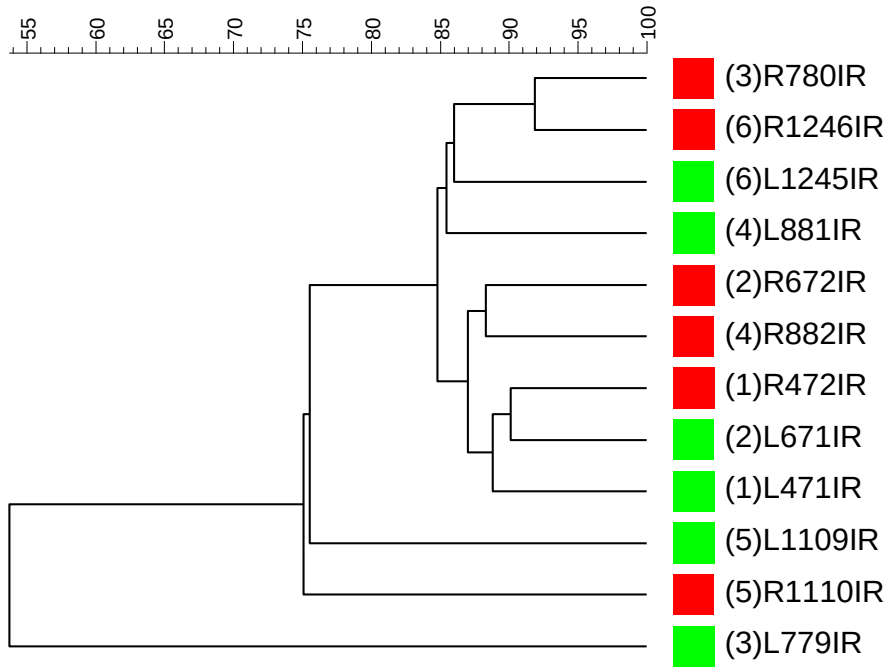
Dendrogram of sheep A4 milk samples



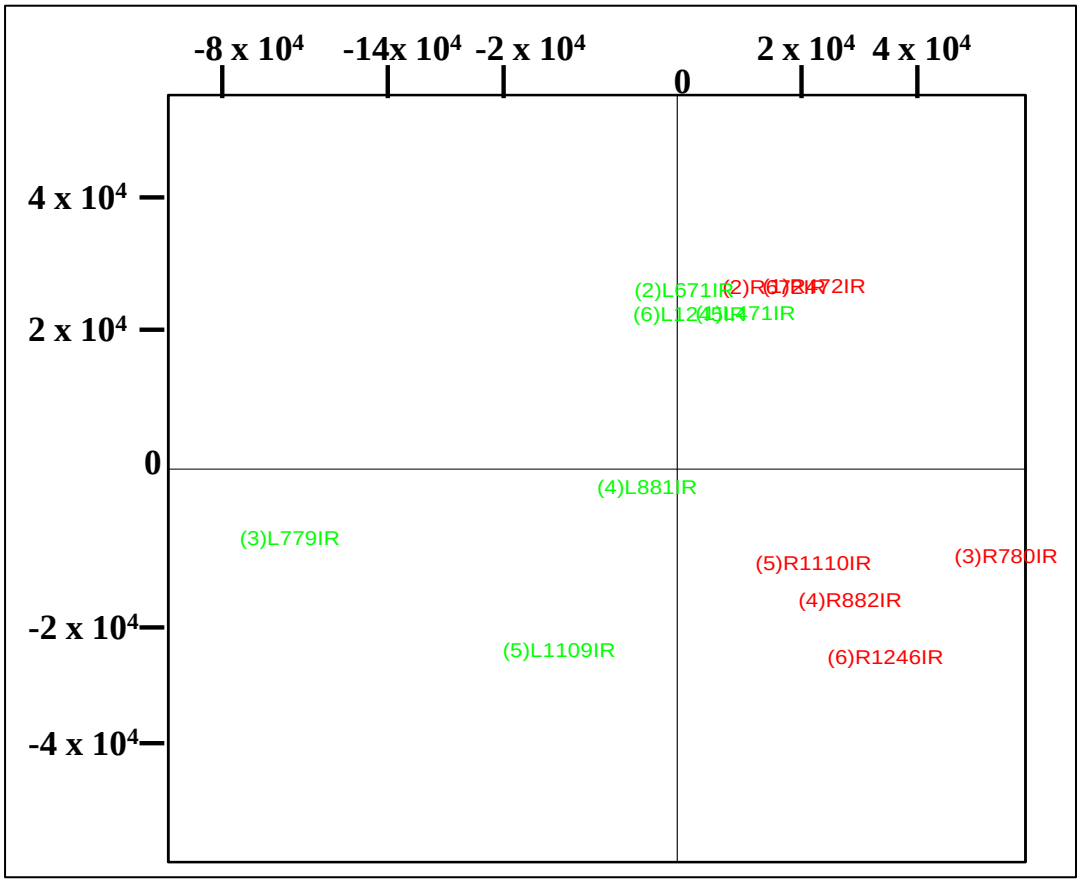
PCA plot of sheep A4 milk samples



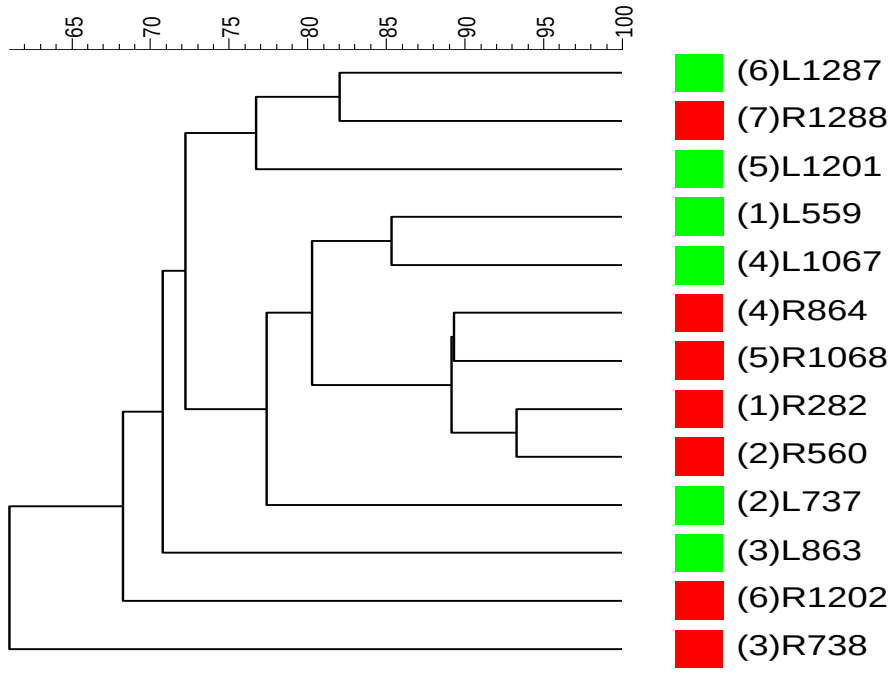
Dendrogram of sheep A41 milk samples



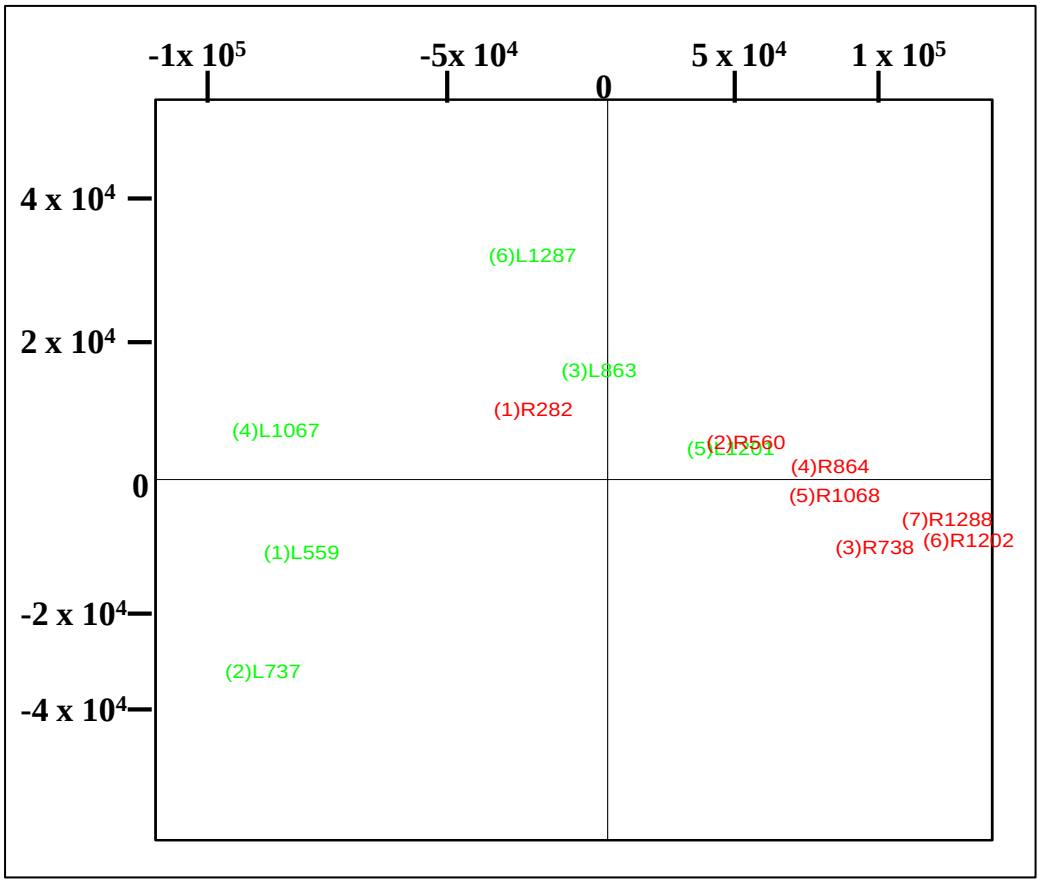
PCA plot of sheep A41 milk samples



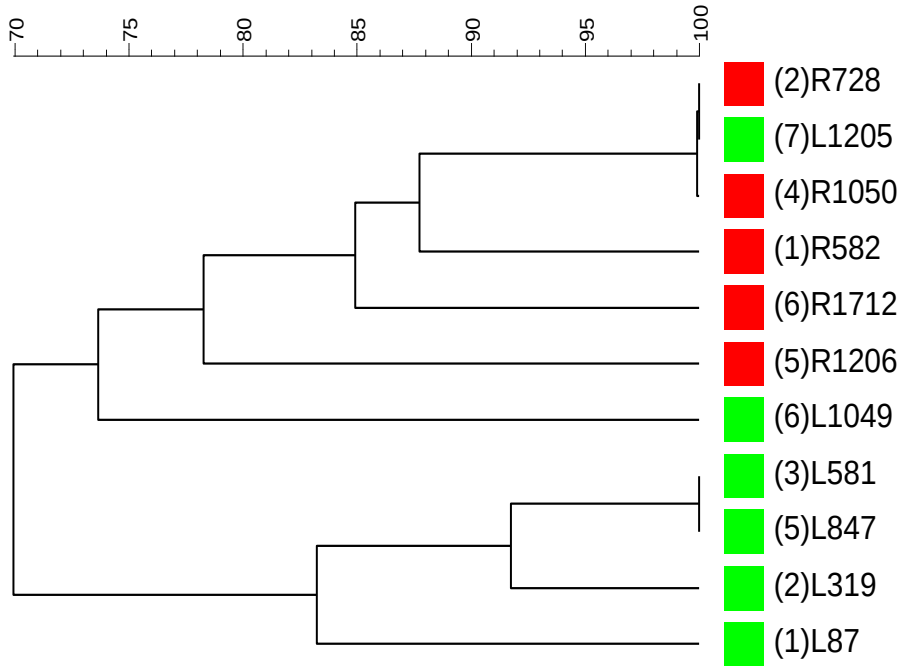
Dendrogram of sheep A44 milk samples



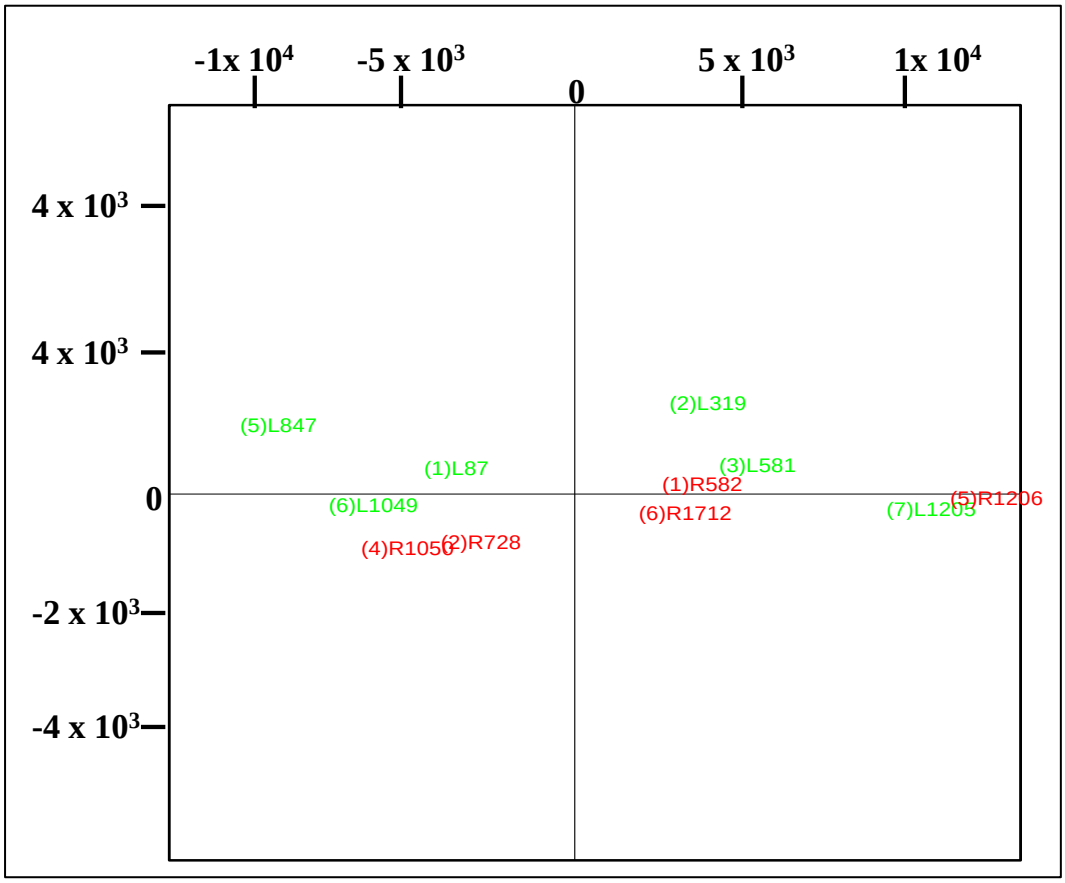
PCA plot of sheep A44 milk samples



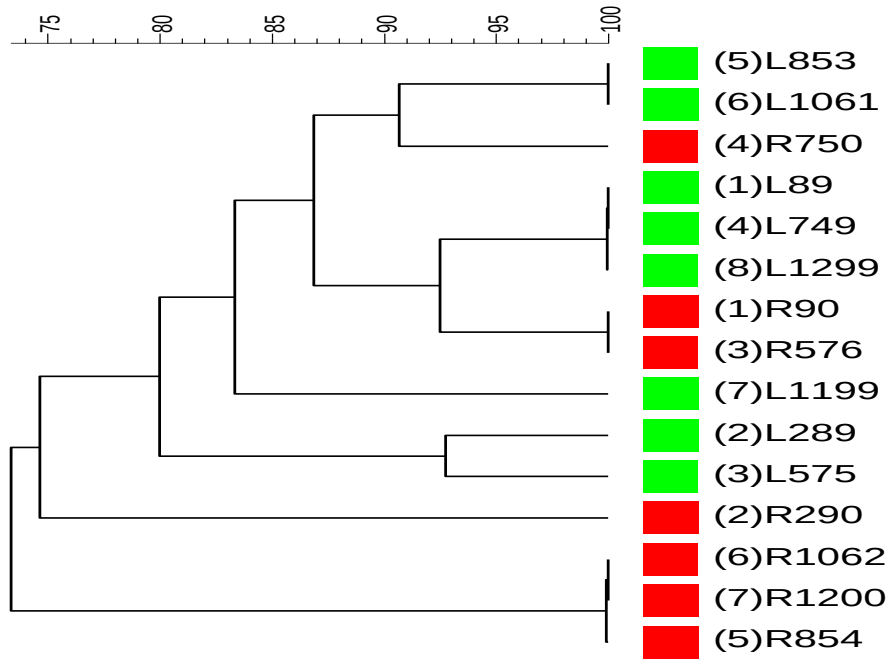
Dendrogram of sheep A45 milk samples



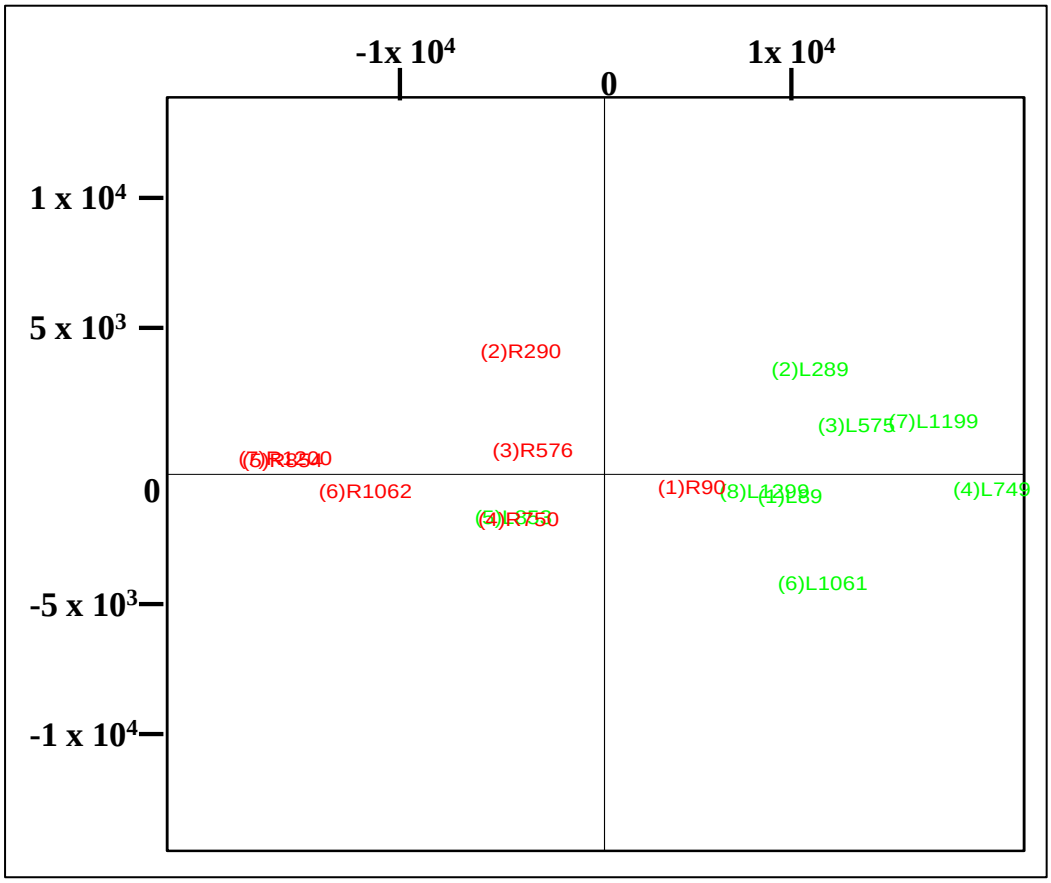
PCA plot of sheep A45 milk samples



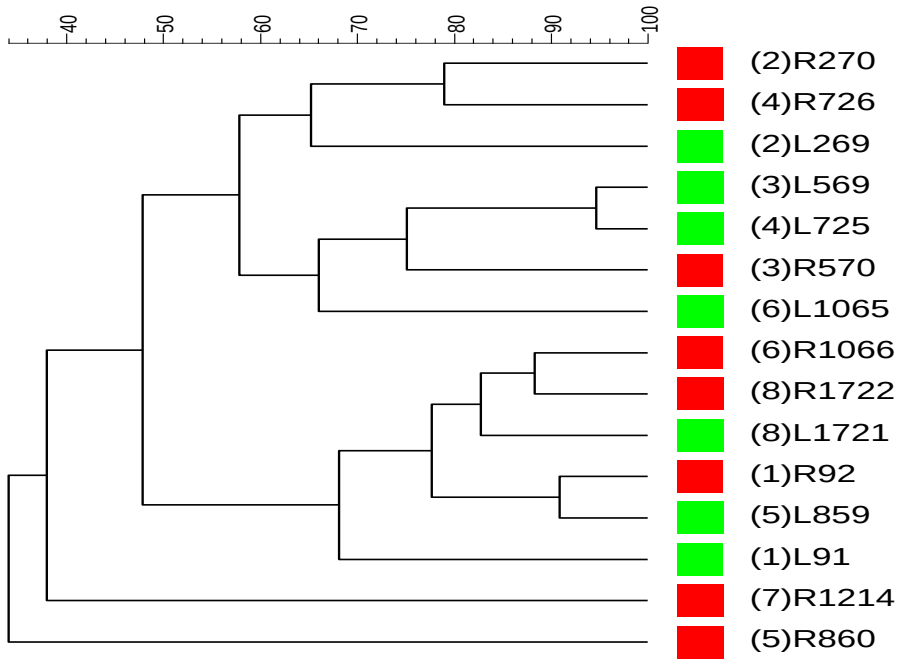
Dendrogram of sheep A46 milk samples



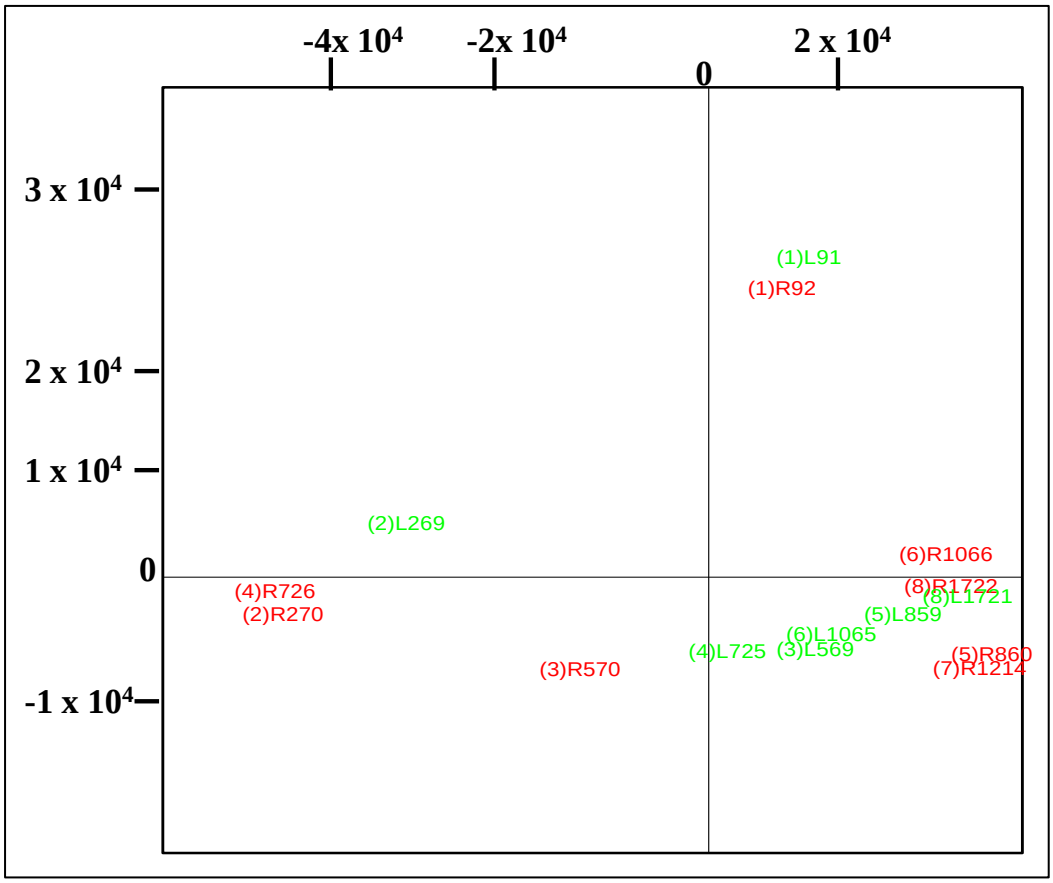
PCA plot of sheep A46 milk samples



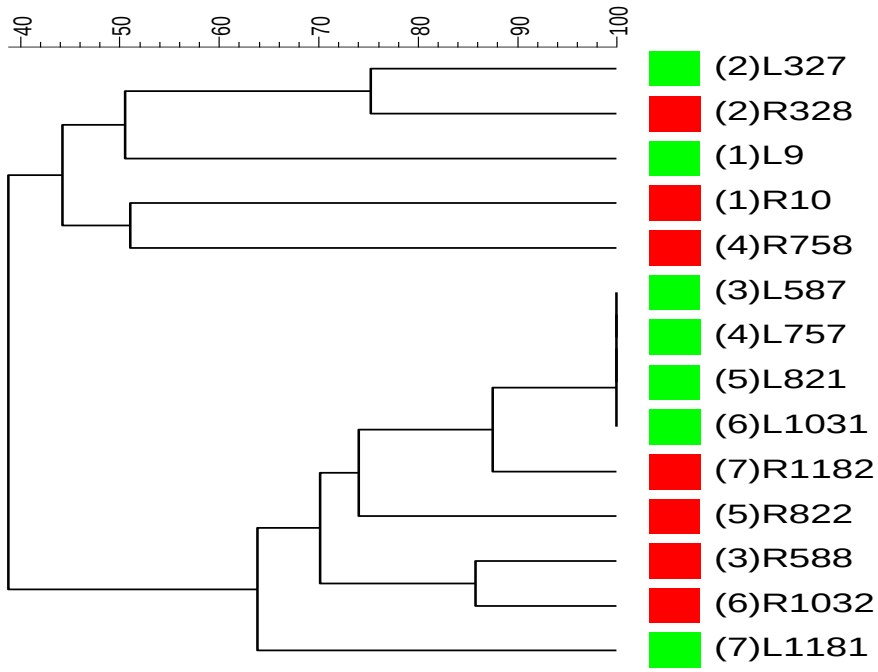
Dendrogram of sheep A47 milk samples



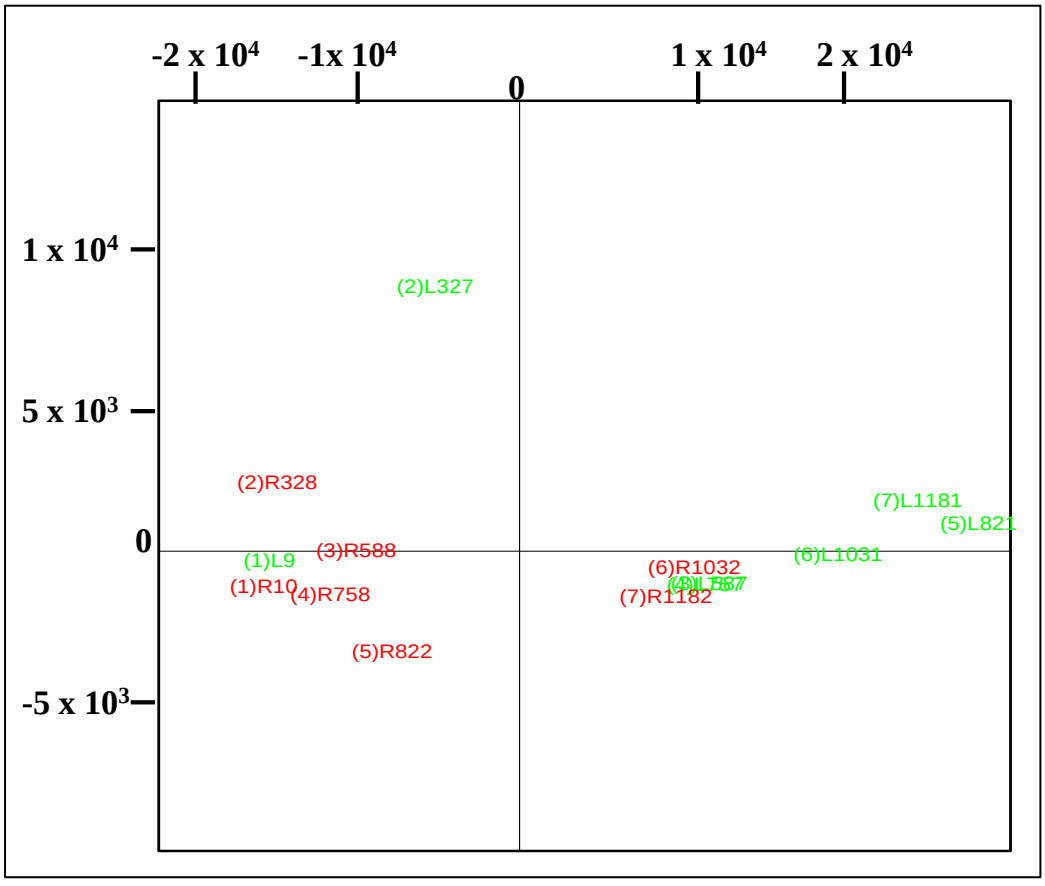
PCA plot of sheep A47 milk samples



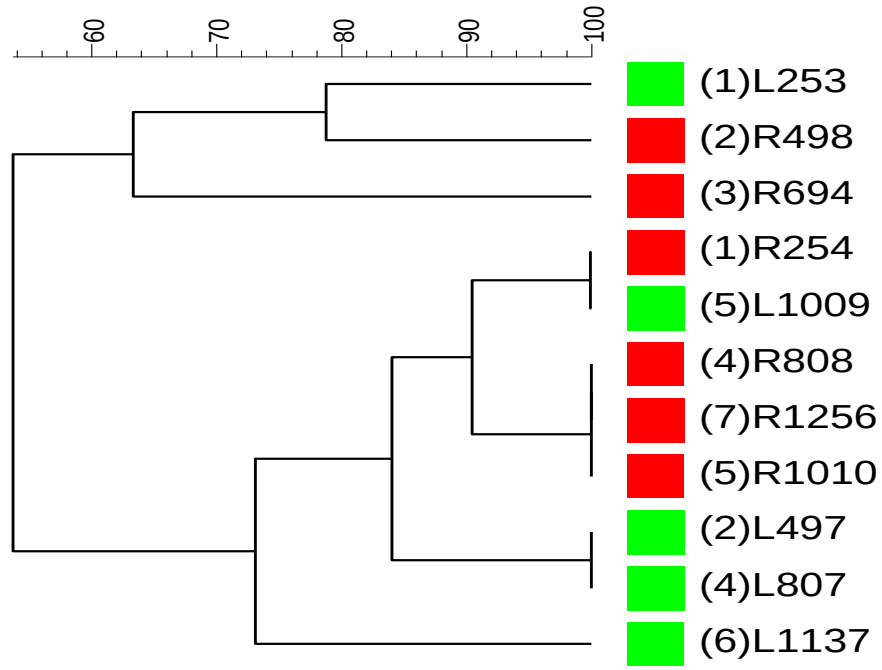
Dendrogram of sheep A5 milk samples



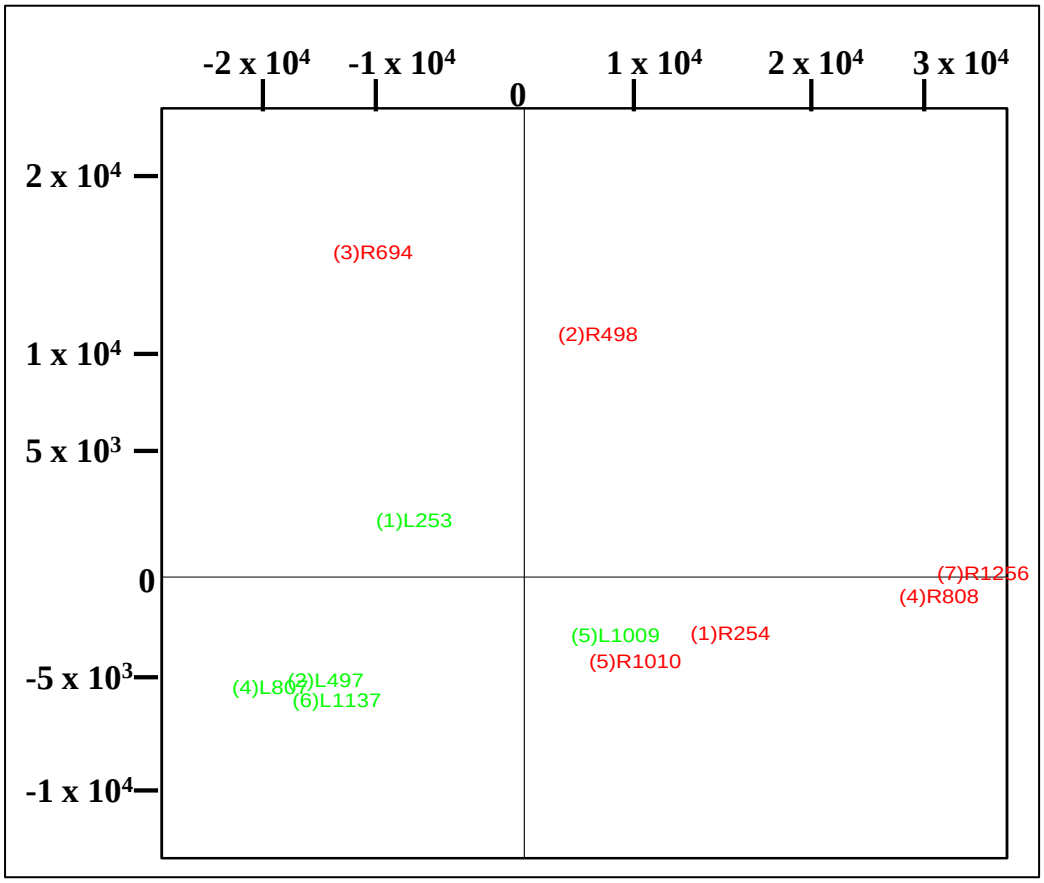
PCA plot of sheep A5 milk samples



Dendrogram of sheep A9 milk samples

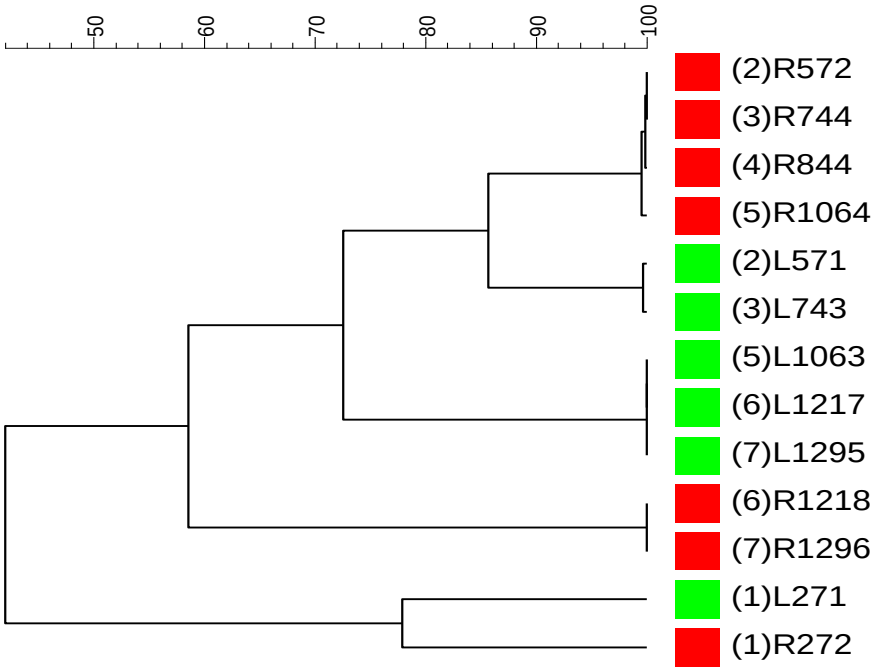


PCA plot of sheep A9 milk samples

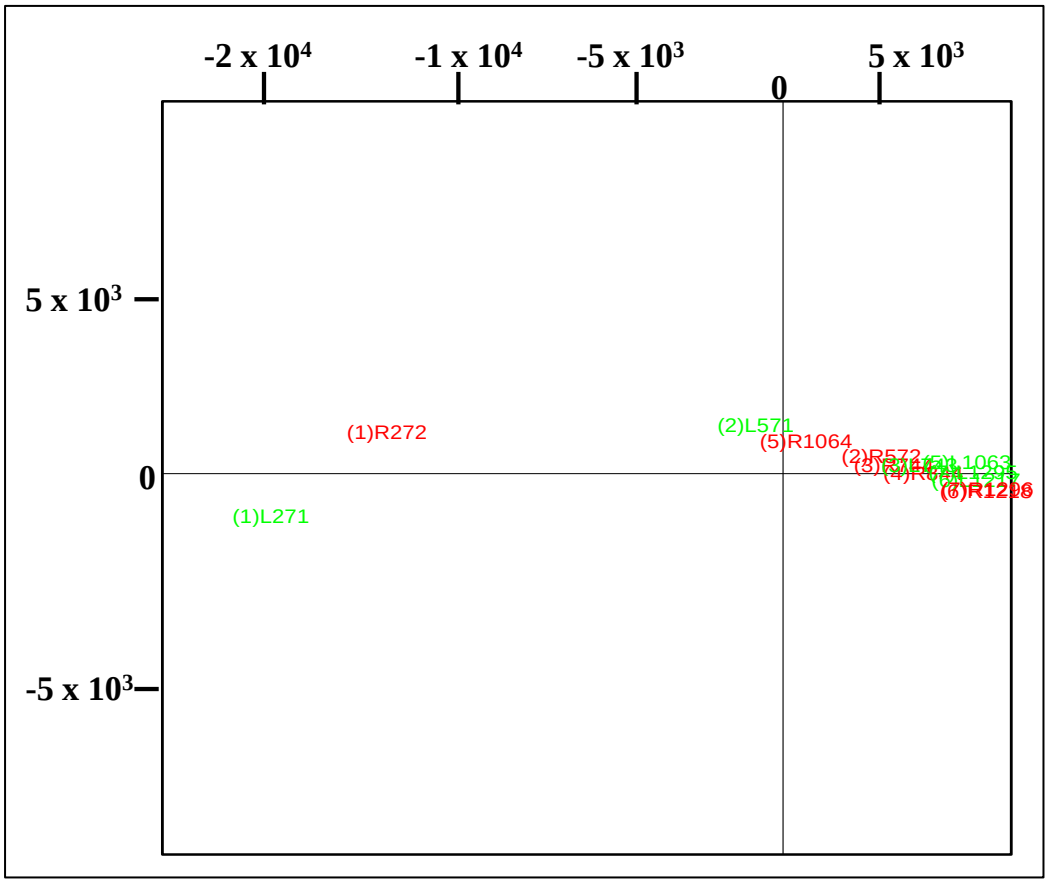


Parity 3 sheep

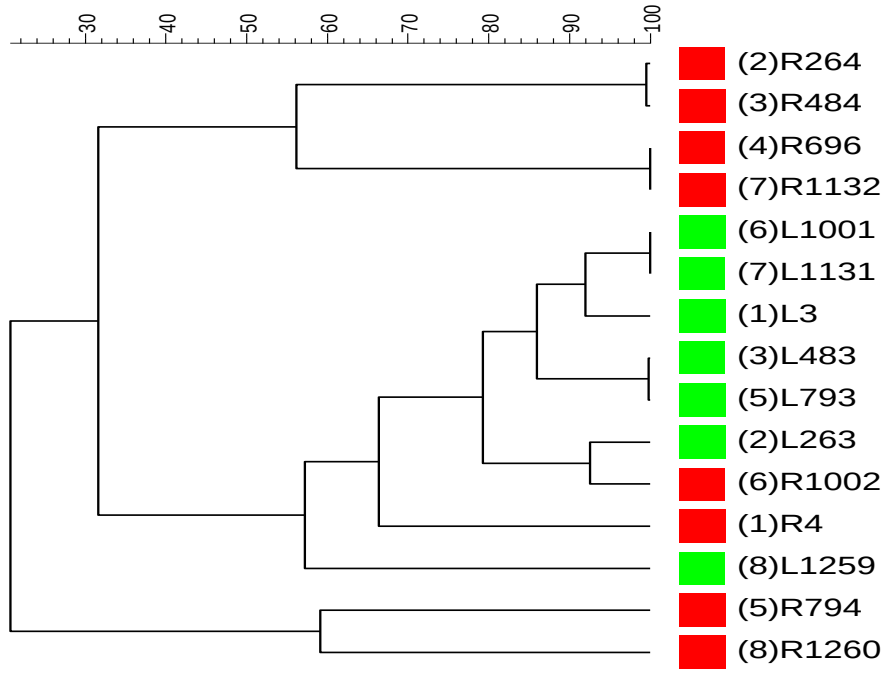
Dendrogram of sheep A12 milk samples



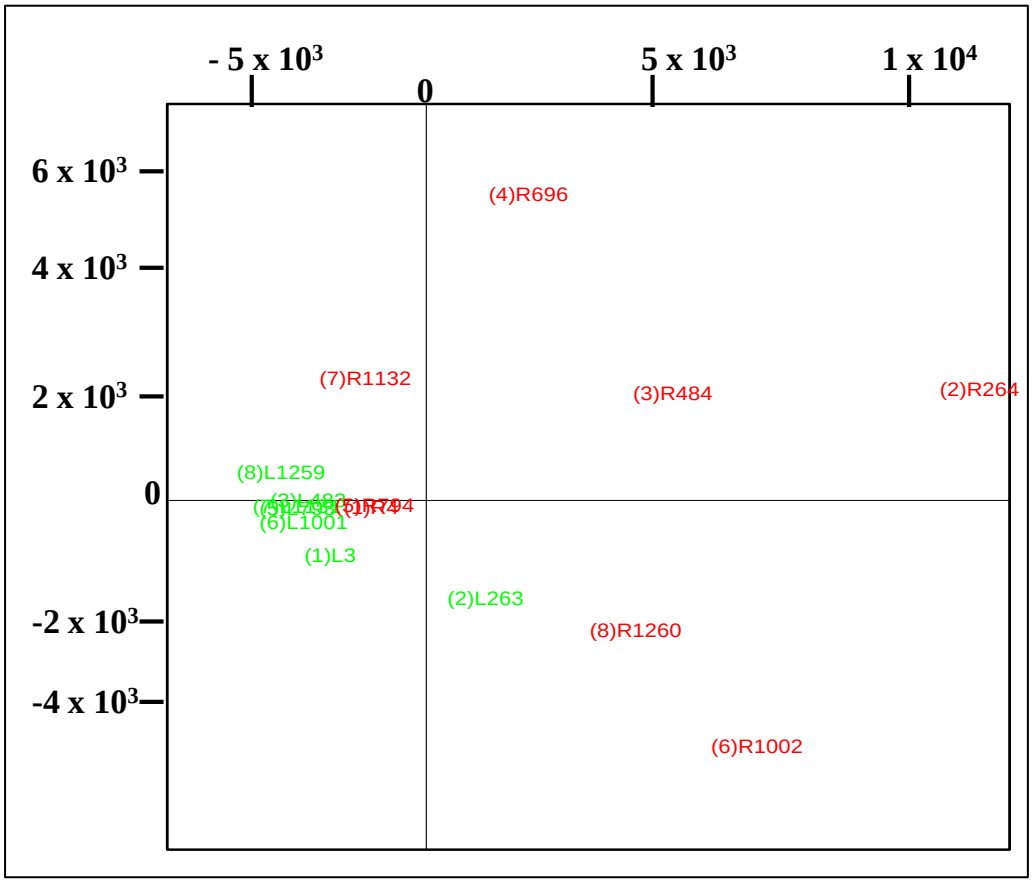
PCA plot of sheep A12 milk samples



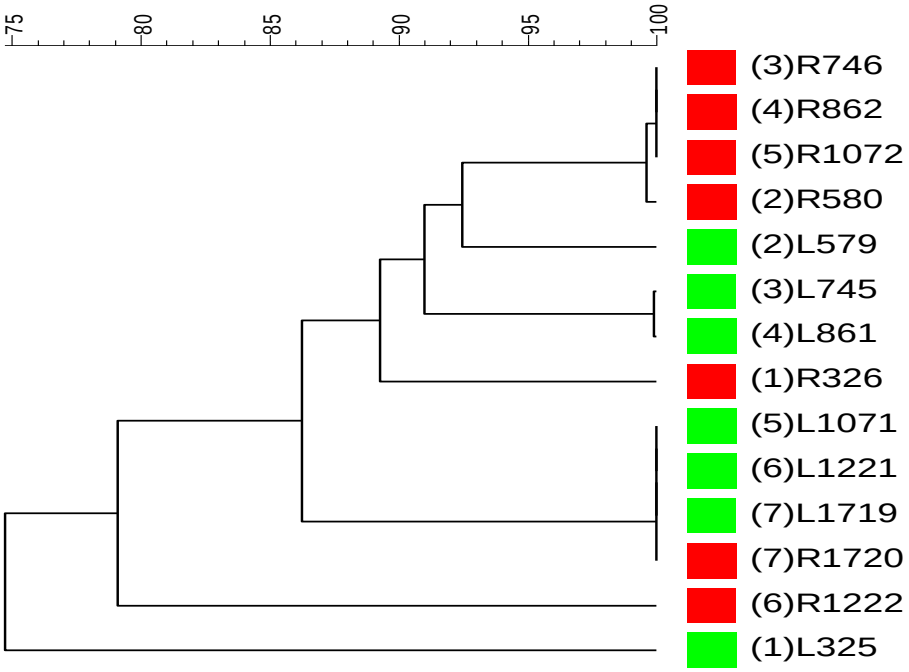
Dendrogram of sheep A2 milk samples



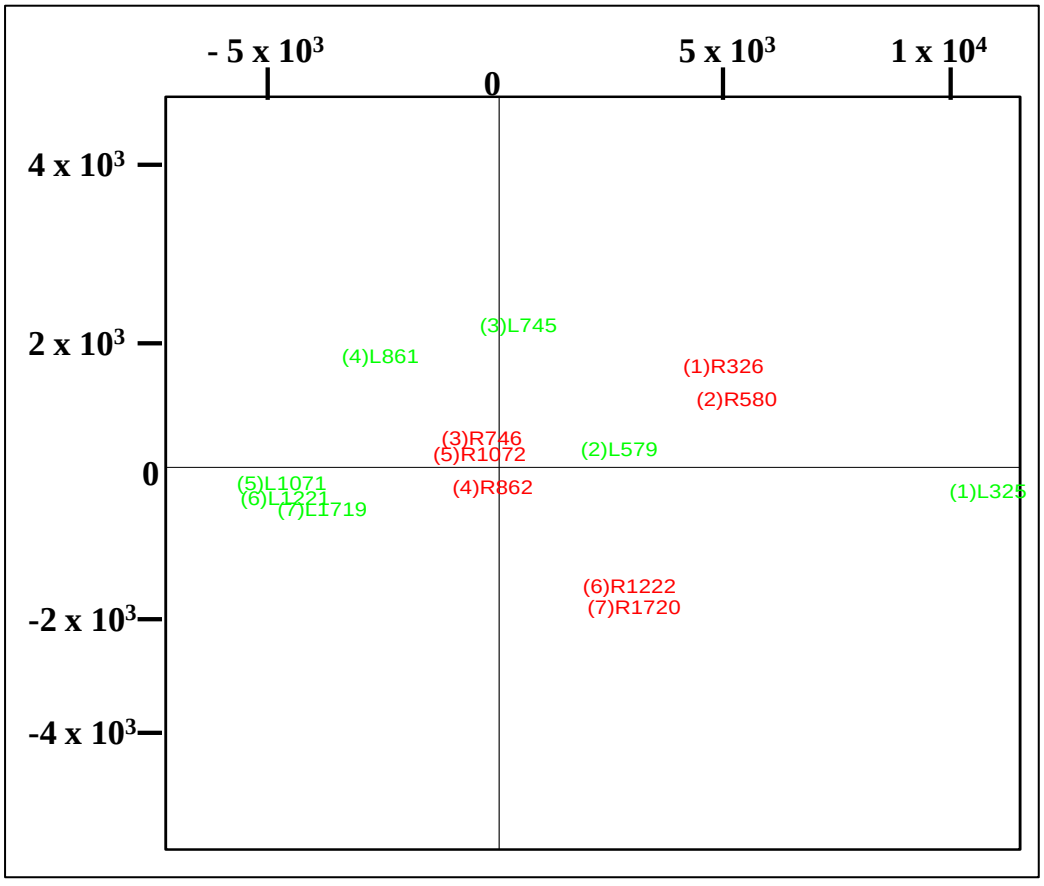
PCA plot of sheep A2 milk samples



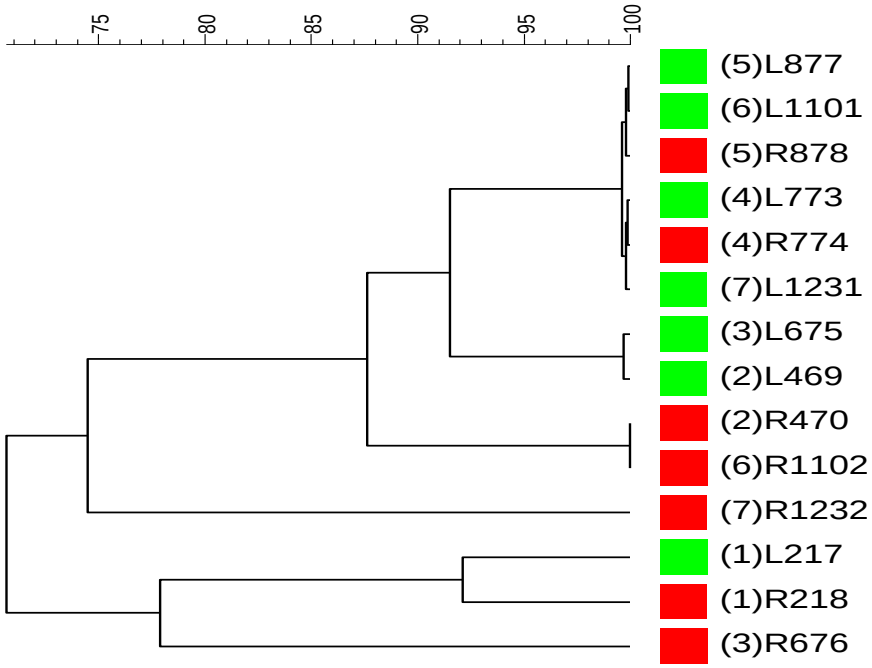
Dendrogram of sheep A22 milk samples



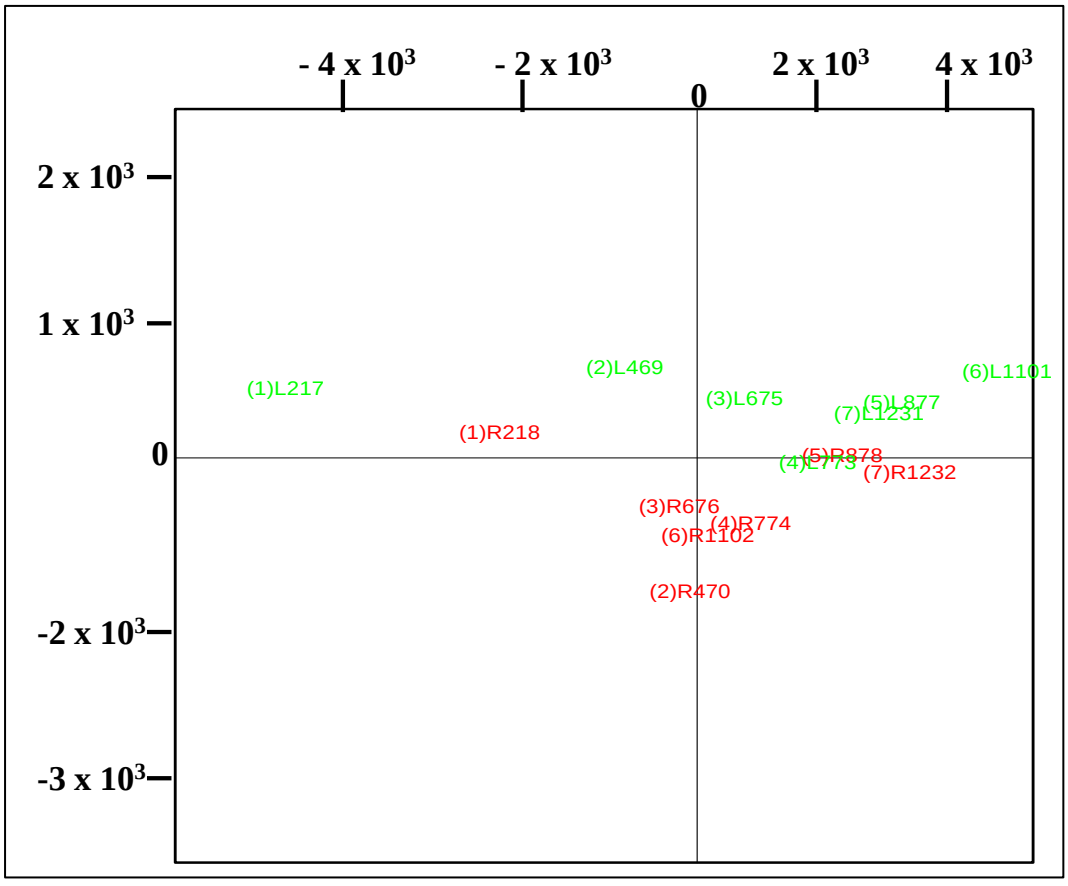
PCA plot of sheep A22 milk samples



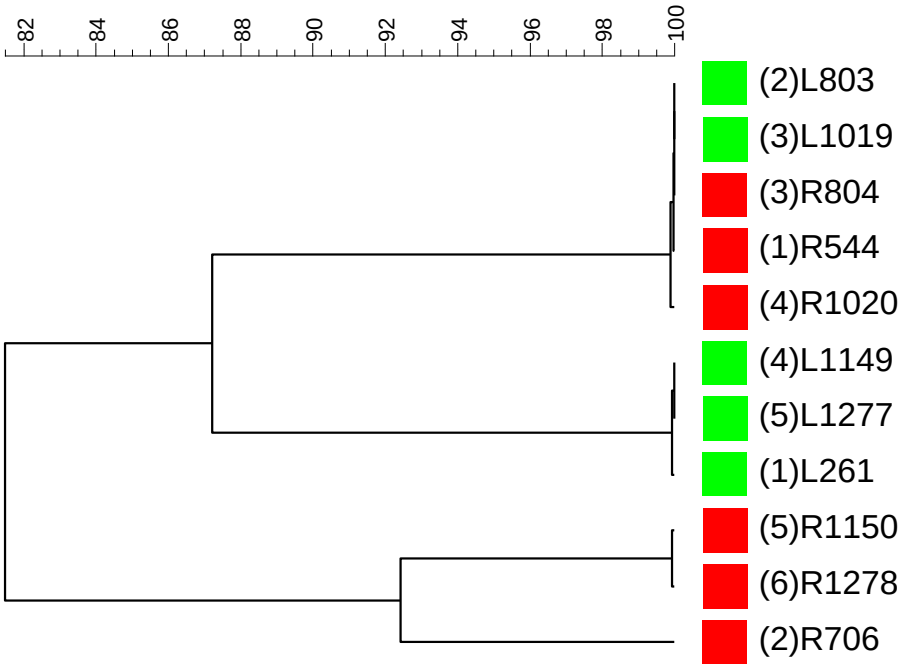
Dendrogram of sheep A29 milk samples



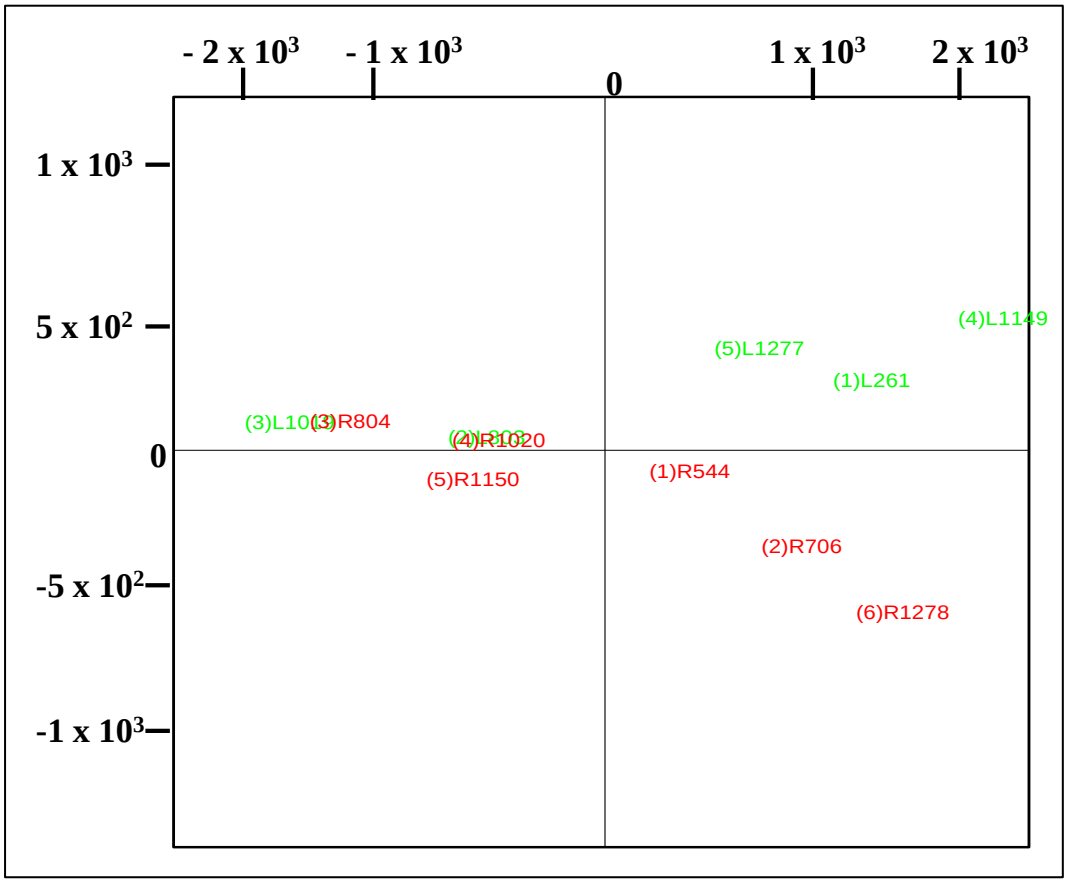
PCA plot of sheep A29 milk



Dendrogram of sheep A3 milk samples

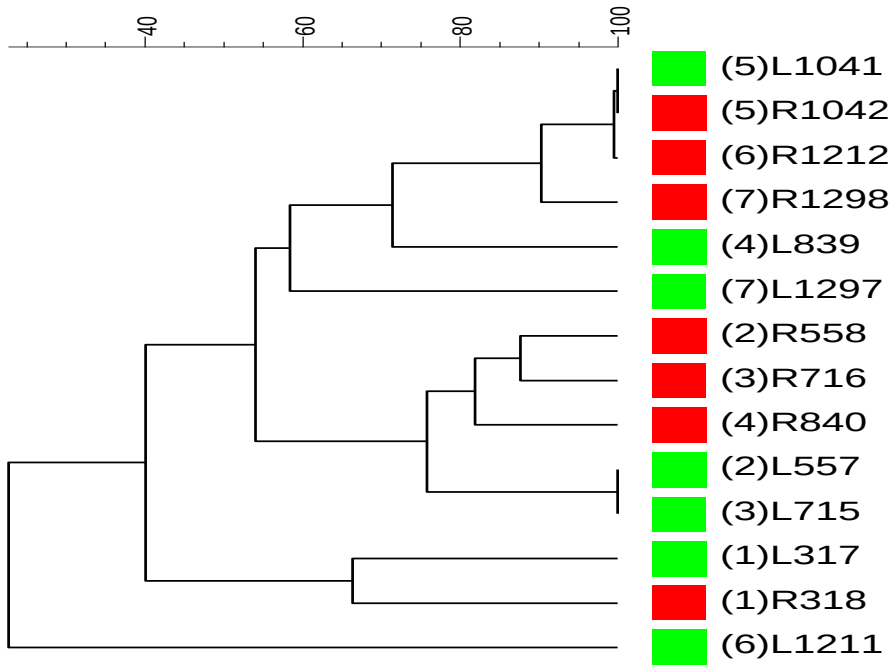


PCA plot of sheep A3 milk samples

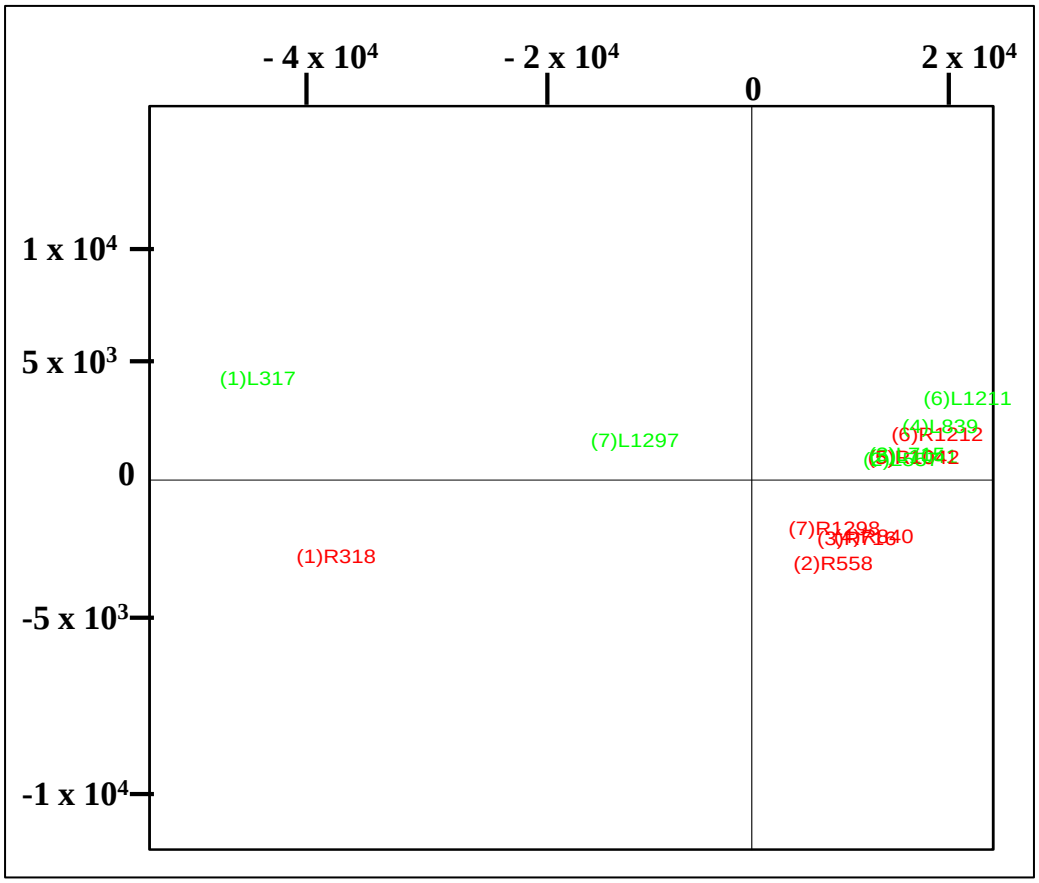


Parity 4 sheep

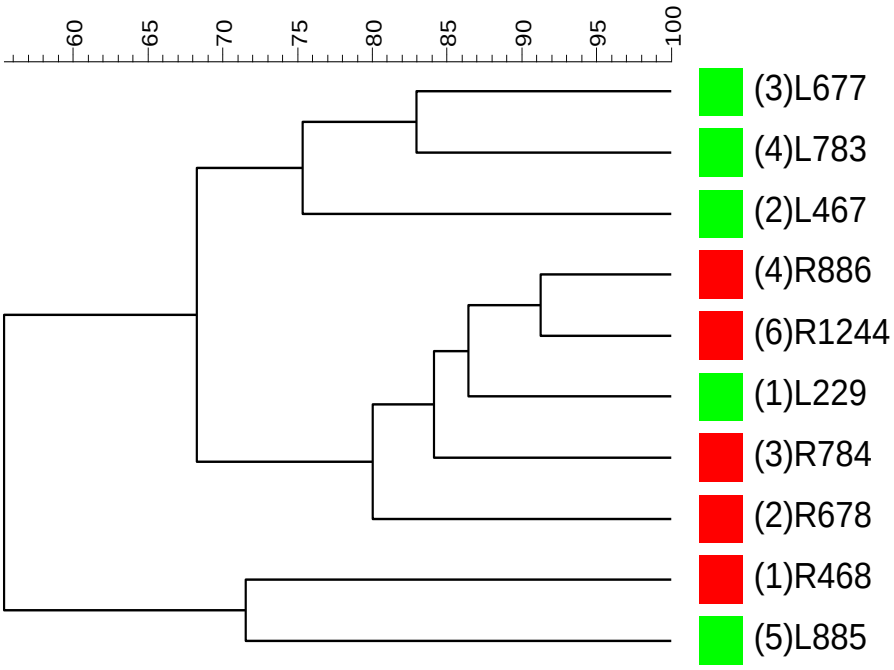
Dendrogram of sheep A16 milk samples



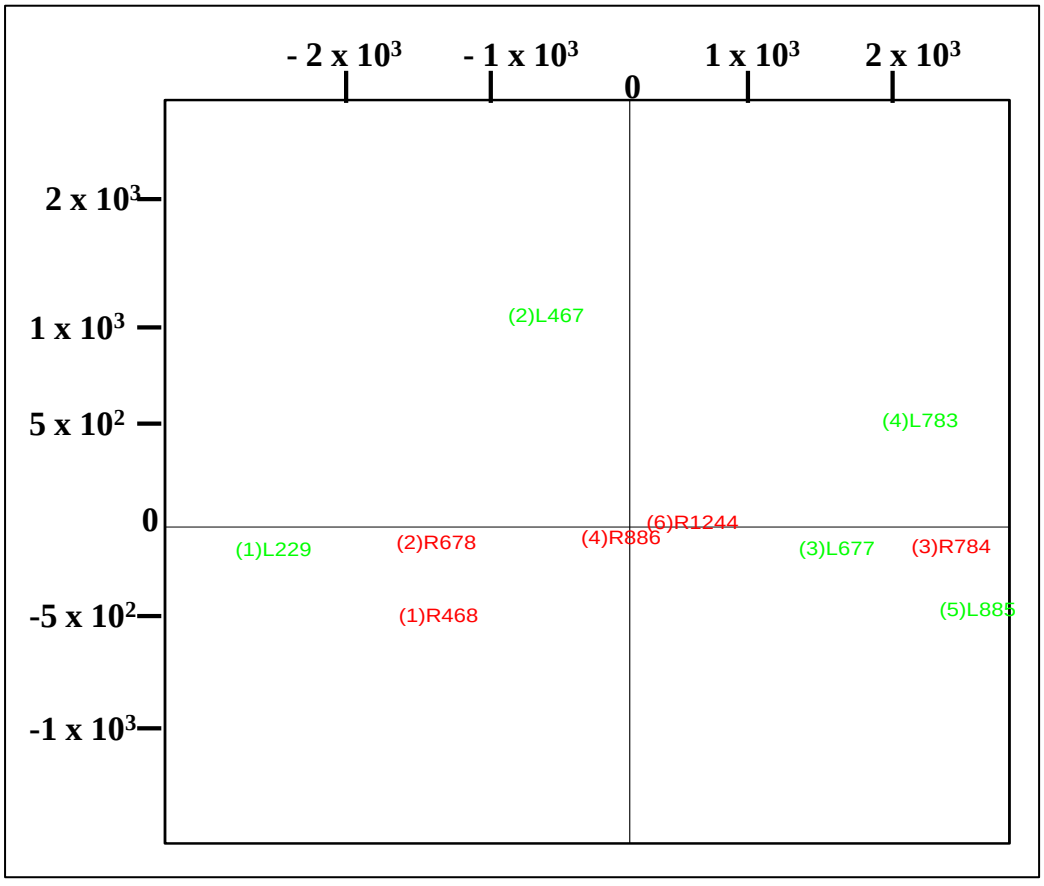
PCA plot of sheep A16 milk samples



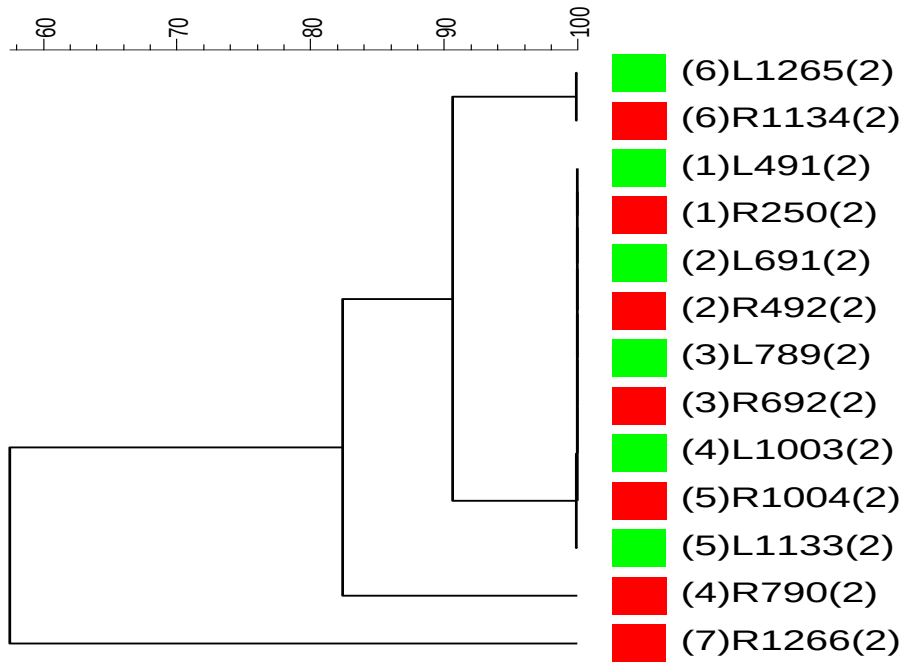
Dendrogram of sheep A24 milk samples



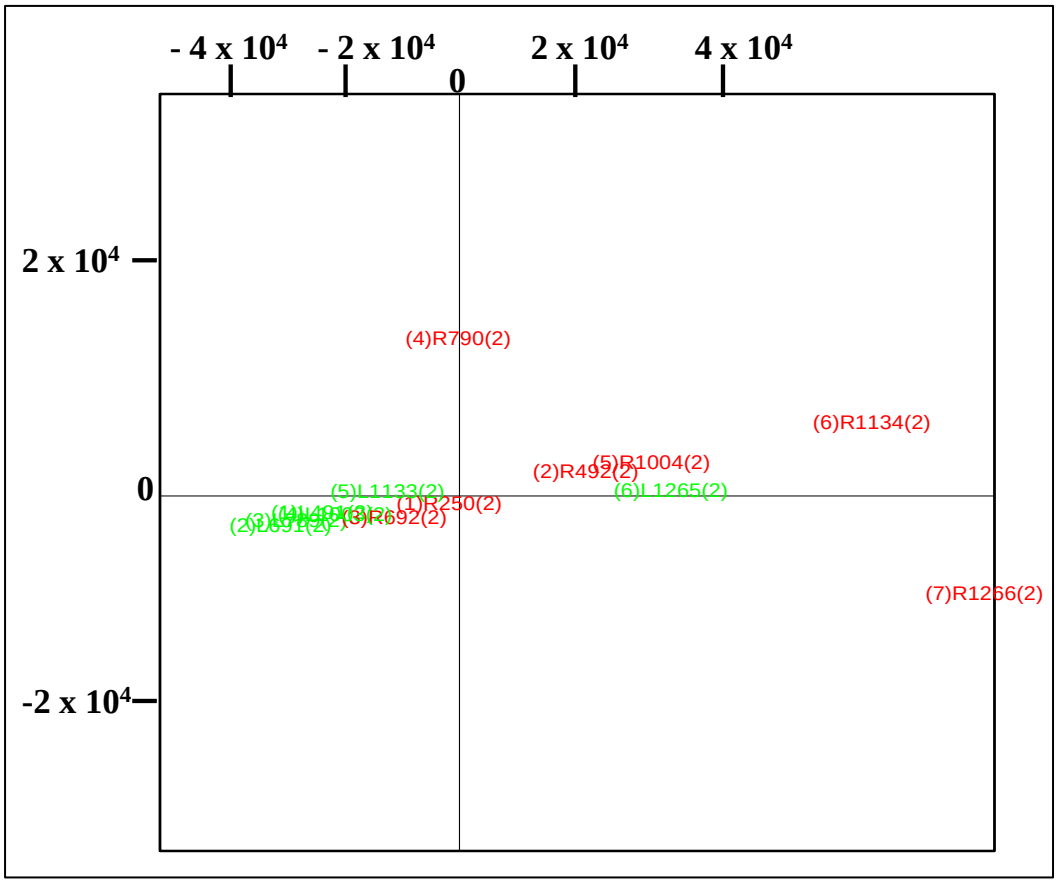
PCA plot of sheep A24 milk samples



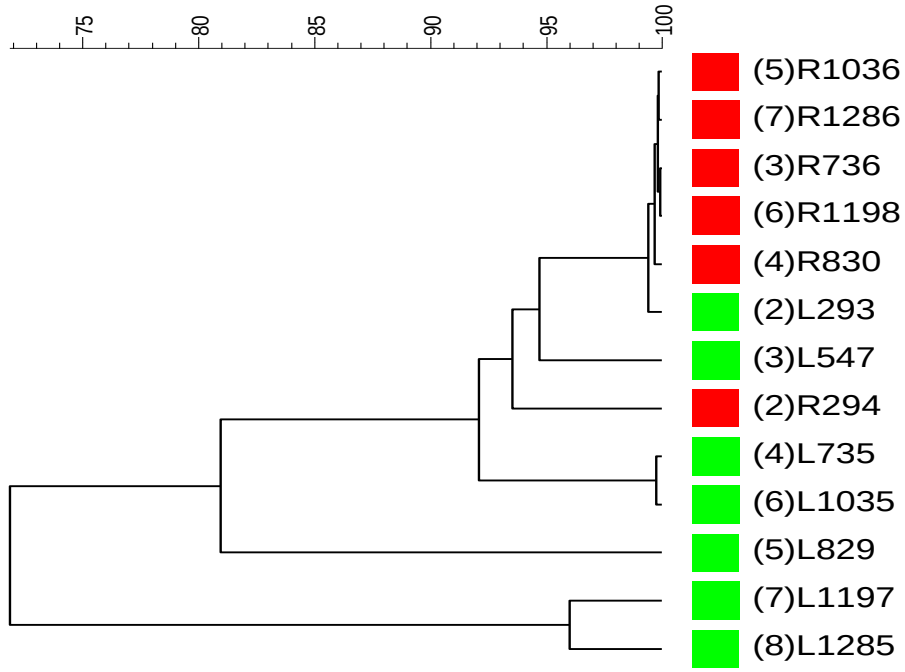
Dendrogram of sheep A40 milk samples



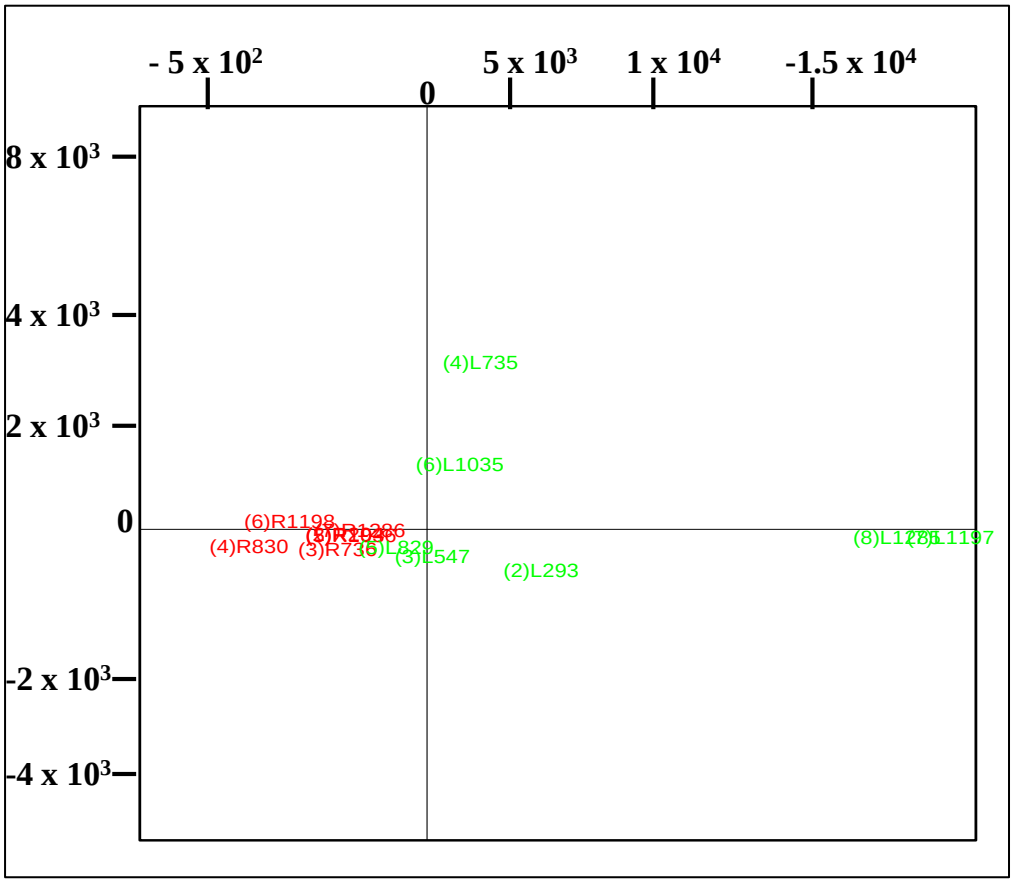
PCA plot of sheep A40 milk samples



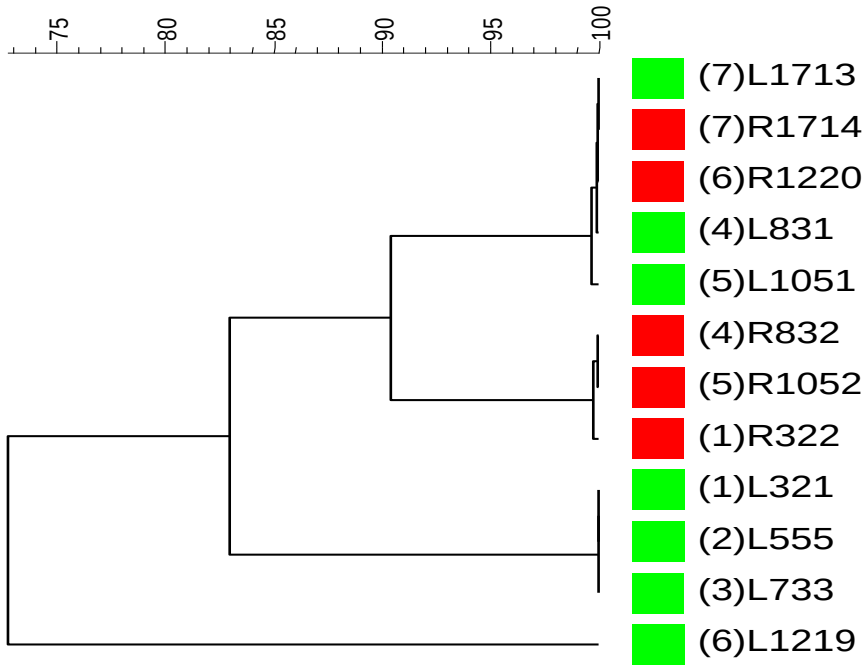
Dendrogram of sheep A43 milk samples



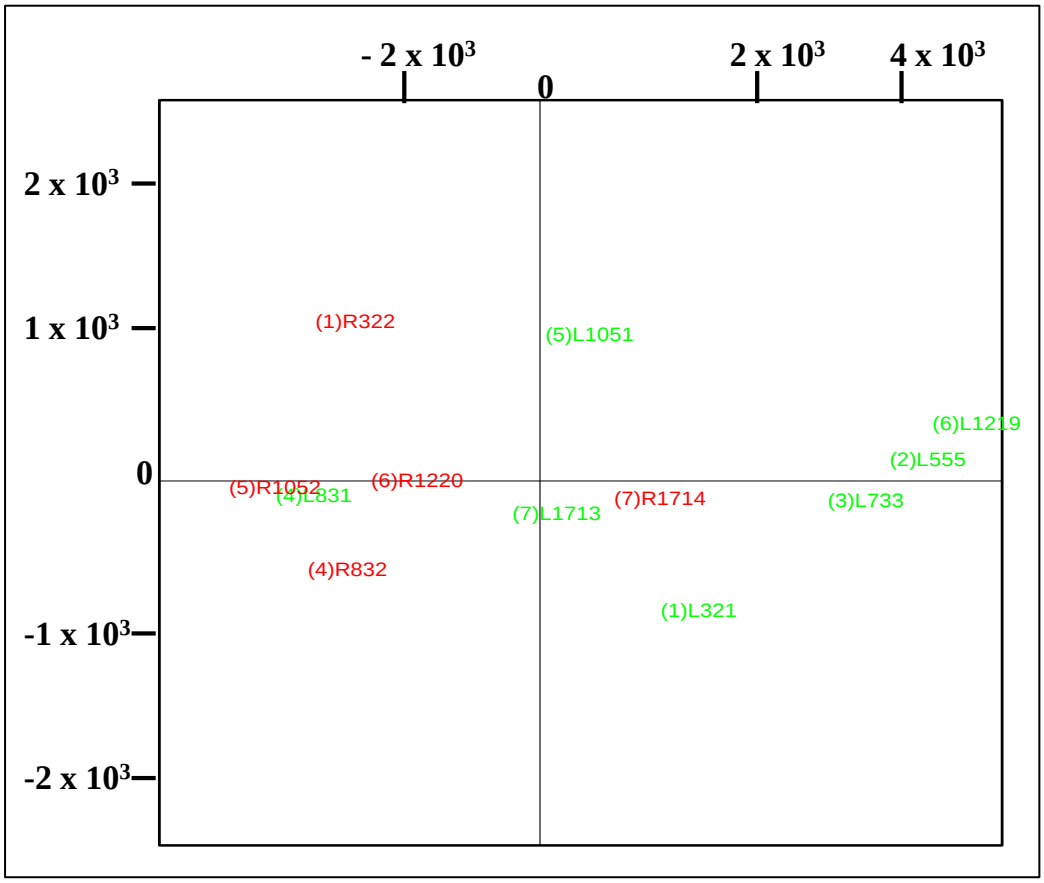
PCA plot of sheep A43 milk samples



Dendrogram of sheep A49 milk samples

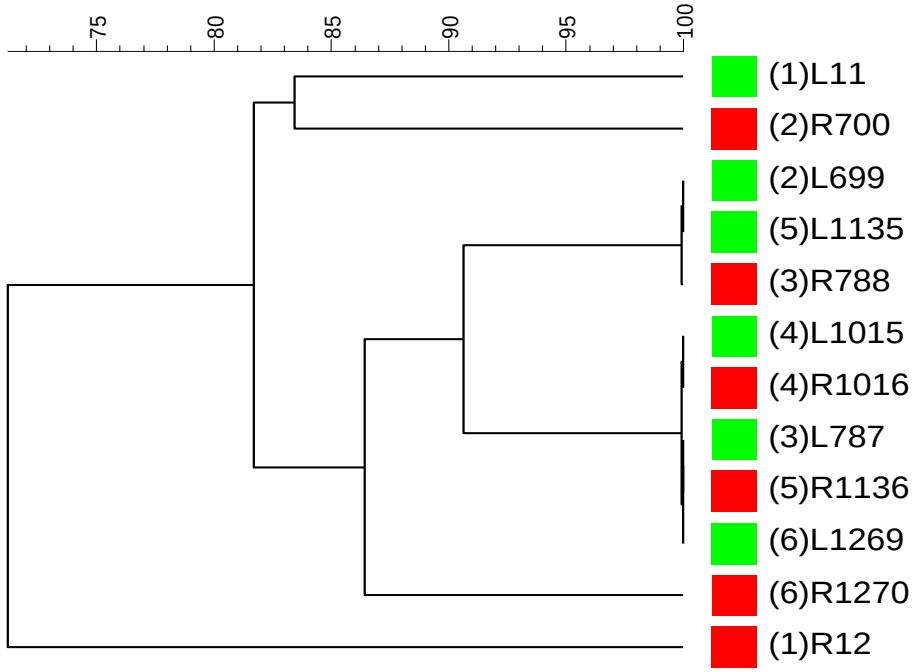


PCA plot of sheep A49 milk samples

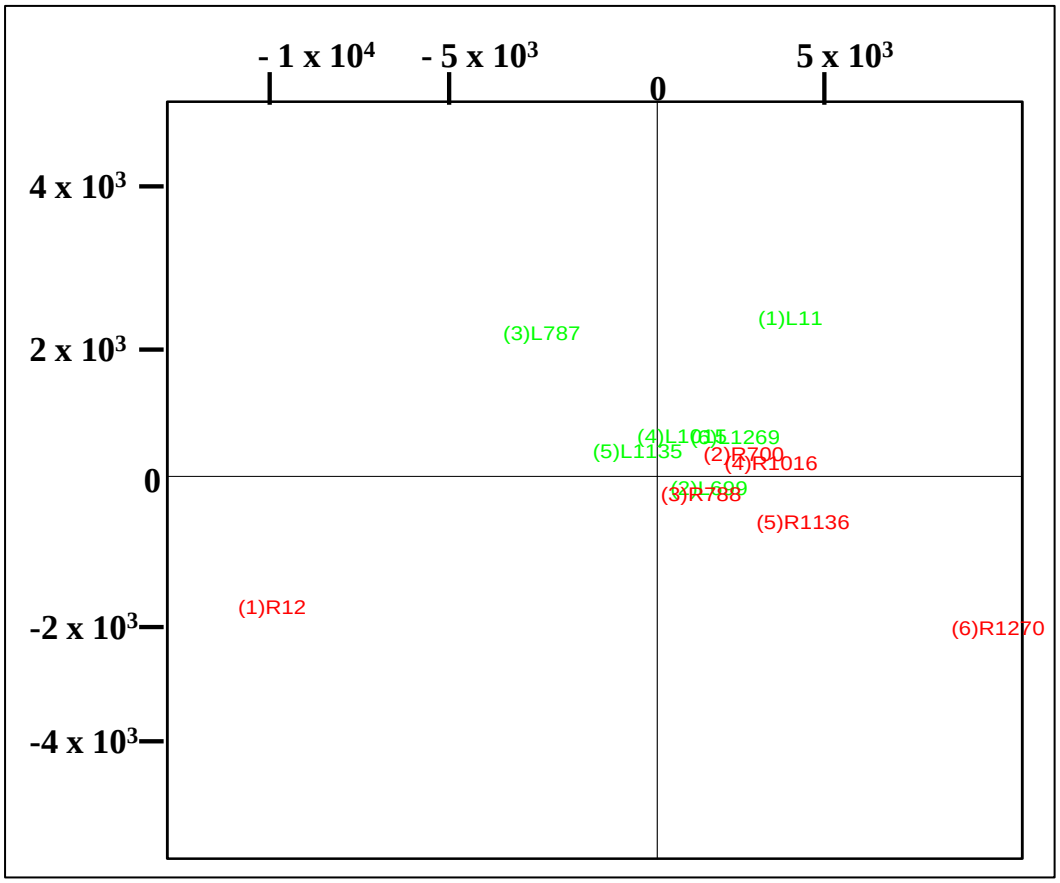


Parity 10 sheep

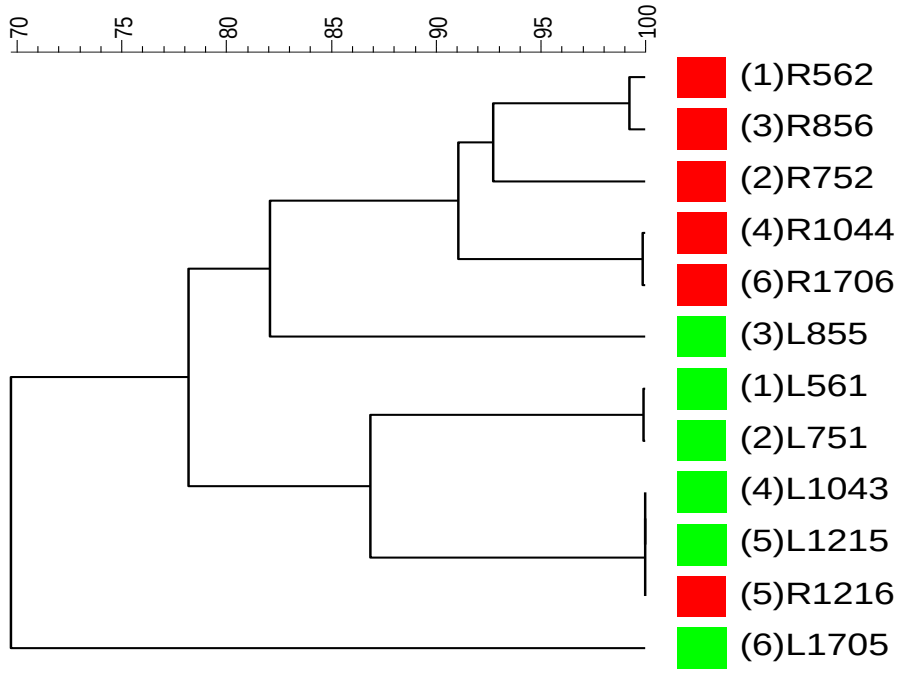
Dendrogram of sheep A6 milk samples



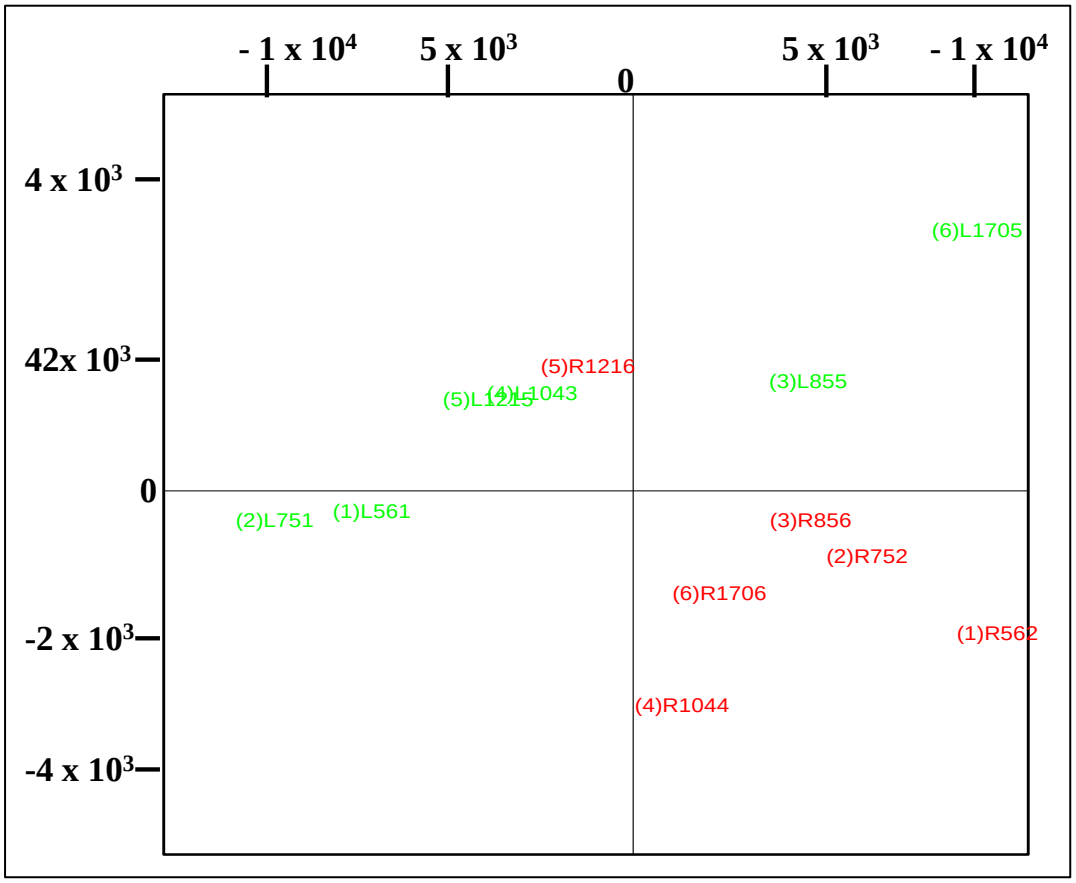
PCA plot of sheep A6 milk samples



Dendrogram of sheep A23 milk samples



PCA plot of sheep A23 milk samples



Appendix 5: MiSeq libraries

Library 1 [42]

1			1	2	3	4	5	6	7	8	9	10	11	12
			N701	**N702**	**N703**	**N704**	**N705**	**N706**	**N707**	**N708**	**N709**	**N710**	**N711**	**N712**
	Sequence		**TCGC CTTA**	**CTAG TACG**	**TTCT GCCT**	**GCTC AGGA**	**AGGA GTCC**	**CATG CCTA**	**GTAG AGAG**	**CCTC TCTG**	**AGCG TAGC**	**CAGC CTCG**	**TGCC TCTT**	**TCCT CTAC**
B	**N502**	**ATAG AGAG**	A45 R1050	A46 R576	A29 L1101	A45 R1050 TR2	A47 L1721 RR1	A15 L27	A41 R266	A47 L1721 RR2	A20 L837	A35 L1277	A46 L289	A43 R736
C	**N503**	**AGAG GATA**	A26 R226	A45 R1712	A22 R1072	A25 R794	A45 R1050 TR1	A24 R696	A16 R1212	A40 R1266	A12 L743	A2 L819	A41 L1109	A20 R740
D	**N504**	**TCTA CTCT**	A12 R1296	A28 L1183	A23 L561	A35 L67	MC2	A28 R826	A162 L557	A26 L875	A16 R840	A26 R670	A23 R1706	A16 R318
E	**N505**	**CTCC TTAC**	A15 L1047	A43 L1035	A27 L663	A27 R664	A39 L1139	A43 L1197	A47 R1214	A47 L91	A22 R746	A35 R1150	A43 R1198	A21 L1159
F	**N506**	**TATG CAGT**	A49 R1714	-	A47 R726	A16 L1297	A23 L1043	A41 R882	A43 R704	A26 L457	A35 L331	A41 R672	A45 R1206	A29 L877
G	**N507**	**TACT CCTT**	A41 L671	A29 R878	A16 R716	A40 R790	A35 R1186	A25 L721	A15 L553	A47 L1065	A20 R1056	A21L3	A47 L1721	A27 L771
H	**N508**	**AGGC TTAG**	A26 R776	A34 R1020	A12 R844	A12 L1217	A25 L827	A43 R830	A23 L483	Model Community	A23 R856	A22 R264	A47 R92	Model Community

[42] For libraries 1-5, the index identifier and sequence for each N5 and N7 primer is provided. The sample information is detailed: Sheep identification; mammary gland half; sample number. Model Community is the model community control which is discussed in further detail in Section 5.2.1.1. All sample locations were randomised across the five libraries using the random number function in Excel. A sample well marked '-' was empty.

Library 2

2			1	2	3	4	5	6	7	8	9	10	11	12
			N701	**N702**	**N703**	**N704**	**N705**	**N706**	**N707**	**N708**	**N709**	**N710**	**N711**	**N712**
	Sequence		**TCGC CTTA**	**CTAG TACG**	**TTC T GCC T**	**GCTC AGGA**	**AGGA GTCC**	**CATG CCTA**	**GTAG AGAG**	**CCTC TCTG**	**AGCG TAGC**	**CAGC CTCG**	**TGCC TCTT**	**TCCT CTAC**
B	**N502**	**ATAG AGAG**	A24 L885	Model Community	A20 R280	A15 R554	A22 R580	A25 L1045	A44 R1288	A4 R502	A49 L1219 RR2	A25 R566	A25 L1293	A29 L675
C	**N503**	**AGAG GATA**	A23 R1044	A57 R1182	A49 L121 9 RR1	A15 L299	A12 R572	A45 L1127	A35 R1074	Model Community	A26 R1106	A3 R544	A28 L825	A16 L1211
D	**N504**	**TCTA CTCT**	A12 R272	A12 R744	A9 R101 0	A6 R1270	A21 R1160	A45 L319	A15 L717	A41 L471	A9 R1256	A28 L589	A44 R796	-
E	**N505**	**CTCC TTAC**	A4 L795	A39 R1274	A24 L229	A50 R1114	A45 R728	A50 R98	A26 R1002	A50 L681	A49 L1219 RR2	A46 L1299	A46 R750	A40 L491
F	**N506**	**TATG CAGT**	A46 R1200	A45 L1049	A16 L104 1	A5 L1031	A50 L873	A6 R788	A25 R828	A44 R282	A24 L885 TR2	A25 R228	A23 L1705	A24 L885 TR1
G	**N507**	**TACT CCTT**	A40 R1134	A46 L1271	A25 L47	A50 R874	A25 R722	A15 R718	A26 R1238	A16 L715	A28 L1259	A39 R812	A15 R300	A29 R1102
H	**N508**	**AGGC TTAG**	A49 L555	Model Community	A49 L831	A39 L487	A27 L1131	A50 R446	A49 R832	A46 R290	A50 R224	A6 R12	A49 L1051	A46 L1199

Library 3

3			1	2	3	4	5	6	7	8	9	10	11	12
			N701	**N702**	**N703**	**N704**	**N705**	**N706**	**N707**	**N708**	**N709**	**N710**	**N711**	**N712**
	Sequence		**TCG CCT TA**	**CTAG TACG**	**TTCT GCCT**	**GCTC AGGA**	**AGGA GTCC**	**CATG CCTA**	**GTAG AGAG**	**CCTC TCTG**	**AGCG TAGC**	**CAGC CTCG**	**TGCC TCTT**	**TCCT CTAC**
B	**N502**	**ATAG AGAG**	A45 L1205	A22 R862	A29 L1231	A21 L755	A35 L1073	A25 R1190	A28 L761	A24 L467	A16 R558	A50 R1236	A24 R468	A44 R864
C	**N503**	**AGAG GATA**	A22 L263	A35 R760	A22 L1221	A12 L571	A27 R456	A23 R1216	A2 L793	A16 L839	A29 L773	A41 L1245	A22 L579	A35 R824
D	**N504**	**TCTA CTCT**	A52 R328	A26 R876	A20 R568	A28 L1075	A15 R842	A45 L1205 TR2	A47 R860	A4 L703	A25 R1294	A49 R1052	A50 L1235	A43 R294
E	**N505**	**CTCC TTAC**	A22 L1719	A40 R492	-	A39 R488	A41 R472	A39 R1140	A27 L1241	A39 L811	A23 R562	A23 L1215	A28 R1260	A43 R1036
F	**N506**	**TATG CAGT**	A26 L225	Model Community	A24 R678	A50 R1236 RR1	A45 L1205 TR1	A39 R76	A47 L725	A28 R334	A50 R1236 RR2	A35 L823	A20 L739	A35 L591
G	**N507**	**TACT CCTT**	A27 R772	A25 L1189	A49 L1713	Model Community	Model Community	A39 R1008	A43 L1285	A27 R1108	A5 L587	A15 R1224	A21 L1033	A15 L1291
H	**N508**	**AGGC TTAG**	A44 L863	A5 L757	A44 L1067	A3 L261	A2 R4	A6 L787	A40 L691	A27 R1132	A47 L269	A29 R774	A44 R1068	A24 L783

Library 4

4			1	2	3	4	5	6	7	8	9	10	11	12
			N701	**N702**	**N703**	**N704**	**N705**	**N706**	**N707**	**N708**	**N709**	**N710**	**N711**	**N712**
	Sequence		**TCGC CTTA**	**CTAG TACG**	**TTCT GCCT**	**GCTC AGGA**	**AGGA GTCC**	**CATG CCTA**	**GTAG AGAG**	**CCTC TCTG**	**AGCG TAGC**	**CAGC CTCG**	**TGCC TCTT**	**TCCT CTAC**
B	**N502**	**ATAG AGAG**	A6 L1015	A22 L1071	A28 R590	A9 R808	A39 L1273	Model Community	A29 R1232	A16 R1298	A9 L1137	A6 L699	A46 6 R1062	A22 L861
C	**N503**	**AGAG GATA**	A41 R780	A5 R1032	A46 L1061	A24 R784	A5 R758	A25 L287	A43 R1286	A6 R1136	A29 R218	A40 L1003	A28 R1184	A27 L1107
D	**N504**	**TCTA CTCT**	A3 L803	A39 L75	A12 L1063	A47 L859	A9 L253	A23 L751 RR2	A24 R1244	A9 R694	A3 R1278	A44 R738	A23 R752	A28 R1076
E	**N505**	**CTCC TTAC**	A46 L575	A40 L1265	A50 L1113	A27 L455	A46 R90	A5 R822	A5 L1181	A49 L321	A5 L327	A4 L265	A3 L1149	A39 R698
F	**N506**	**TATG CAGT**	A15 L1223	A21 L585	A44 R1202	A21 R756	A6 L1015 TR2	A12 R1064	A49 R322	A44 L737	A15 L841	A263L669	A26 L1001	Model Community
G	**N507**	**TACT CCTT**	A15 R28	A23 L751	A12 L271	A6 L1135	A3 R804	A3 R706	-ve	A9 L497	A23 L751 RR1	A20 R1208	A6 L1015 TR1	Model Community
H	**N508**	**AGGC TTAG**	A21 R330	A5 R588	A22 L745	A43 L547	A24 L677	A29 R676	A5 L9	A47 R1722	A26 R458	A3 L1019	A26 L1105	A23 R484

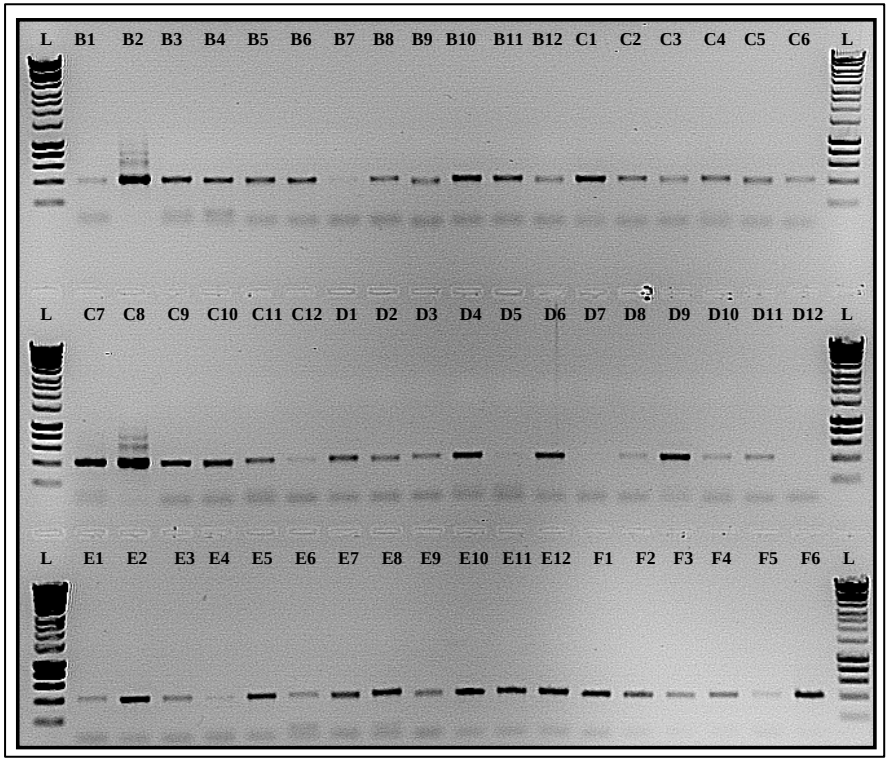
Library 5

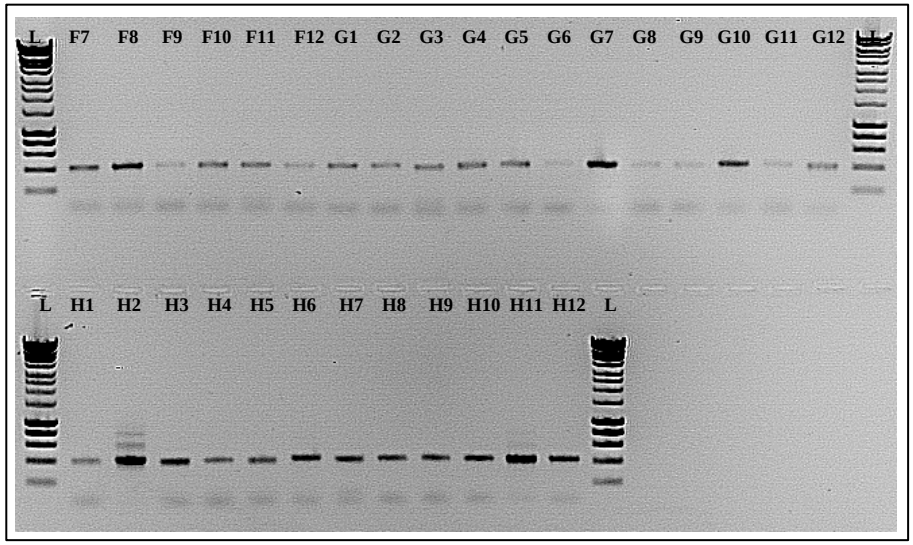
5			1	2	3	4	5	6	7	8	9	10	11	12
			N701	**N702**	**N703**	**N704**	**N705**	**N706**	**N707**	**N708**	**N709**	**N710**	**N711**	**N712**
	Sequence		**TCGC CTTA**	**CTAG TACG**	**TTCT GCCT**	**GCTC AGGA**	**AGGA GTCC**	**CATG CCTA**	**GTAG AGAG**	**CCTC TCTG**	**AGCG TAGC**	**CAGC CTCG**	**TGCC TCTT**	**TCCT CTAC**
B	**N502**	**ATAG AGAG**	A20 L1709	A20 L1207	A12 R1218	A46 L89	A22 R326	A43 L829	A5 R10	A12 L1295	A23 L855	A46 L749	A29 L469	A20 L567
C	**N503**	**AGAG GATA**	A16 L317	A43 L735	A21 L329	A22 L325	A20 L1055	A46 L89 RR2	A20 R838	A40 R692	A45 L87	A20 L1709 TR2	A9 L807	A29 L217
D	**N504**	**TCTA CTCT**	A40 L1133	A47 L569	Model Community	Model Community	A6 R700	A40 L789	A9 R254	A5 L821	A41 R1246	A20 R1710	A43 L293	A6 R1016
E	**N505**	**CTCC TTAC**	A20 L1709 TR1	A24 R886	A45 L581	Model Community	A21 R1034	A44 L559	A45 R582	A50 L465	A50 L97	A45 L847	A35 R332	A21 R586
F	**N506**	**TATG CAGT**	A26 L775	A41 L779	A39 L239	A50 L777	A22 R1720	A46 R854	A47 R270	A49 R1220	A46 L89 RR1	A16 R1042	A35 L1185	A41 L881
G	**N507**	**TACT CCTT**	A9 R498	A46 L853	A39 R240	A39 L1007	A4 L501	A44 L1201	A44 L1287	Model Community	A47 R1066	A40 R1004	A21 R820	A47 R570
H	**N508**	**AGGC TTAG**	-	A29 R470	A6 L1269	A40 R250	A22 R1222	A49 L733	A44 R560	A20 L279	A25 L565	A39 L697	A6 L11	A9 L1009

Appendix 6: PCR results for MiSeq libraries 2-5

Library 2

27F-rd1/338R-rd2 PCR results [43]

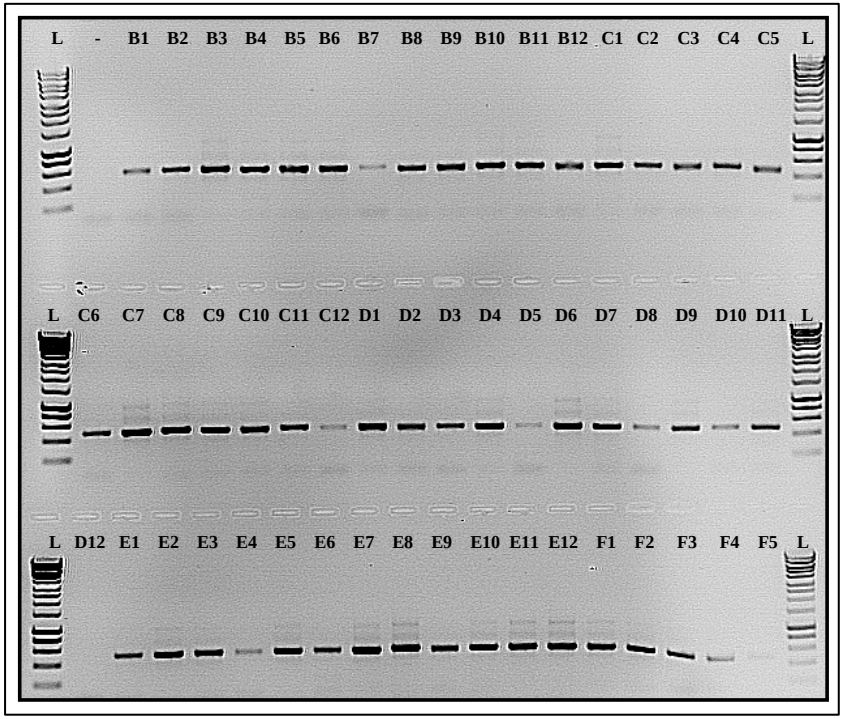


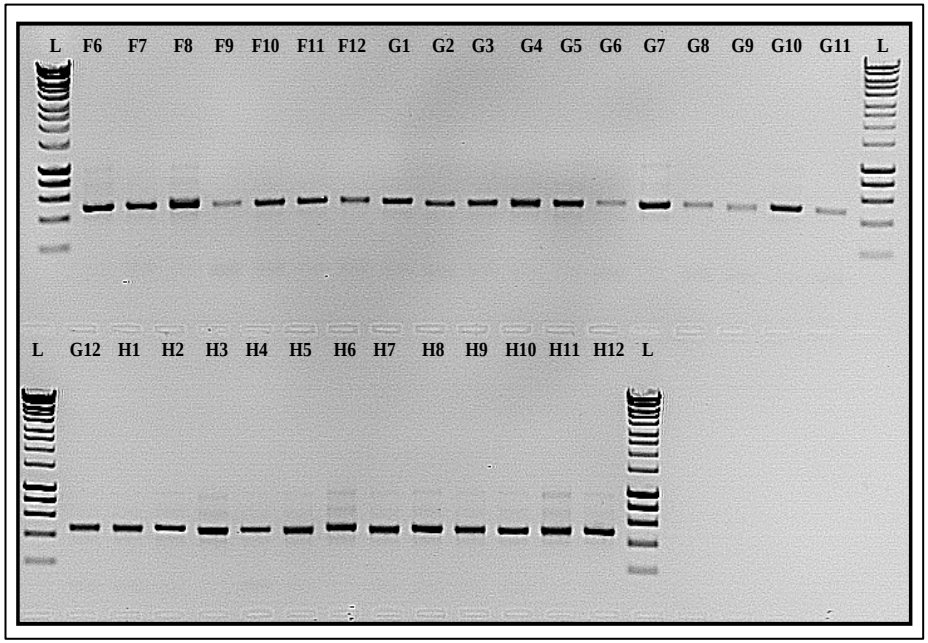


[43] The letter and number for each sample corresponds to the sample identification on the template for library 2 in Appendix 5. Samples with weak amplification were repeated.

Library 2

N5/N7 PCR results [44]

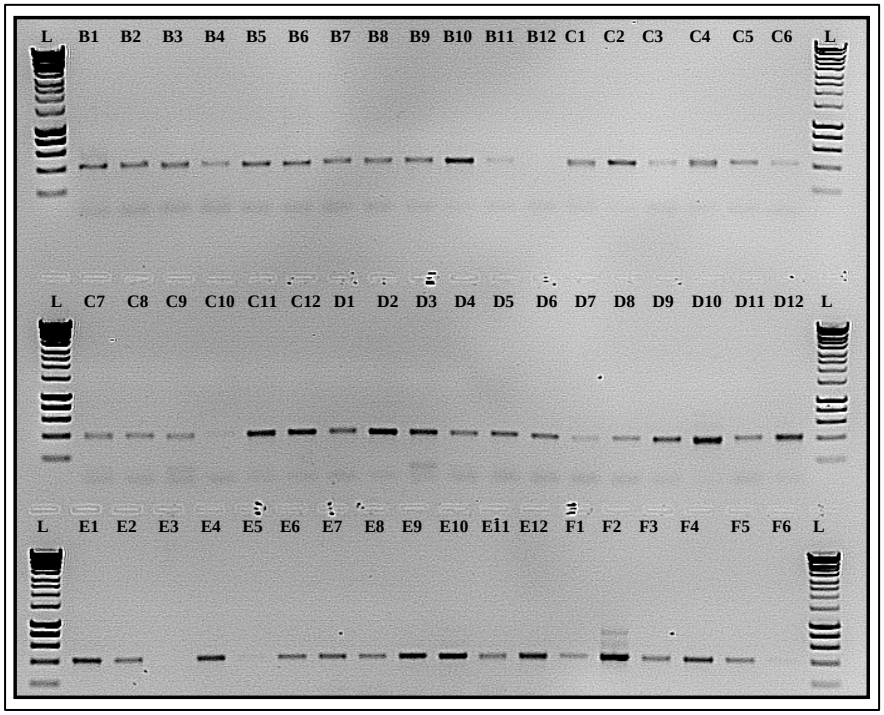


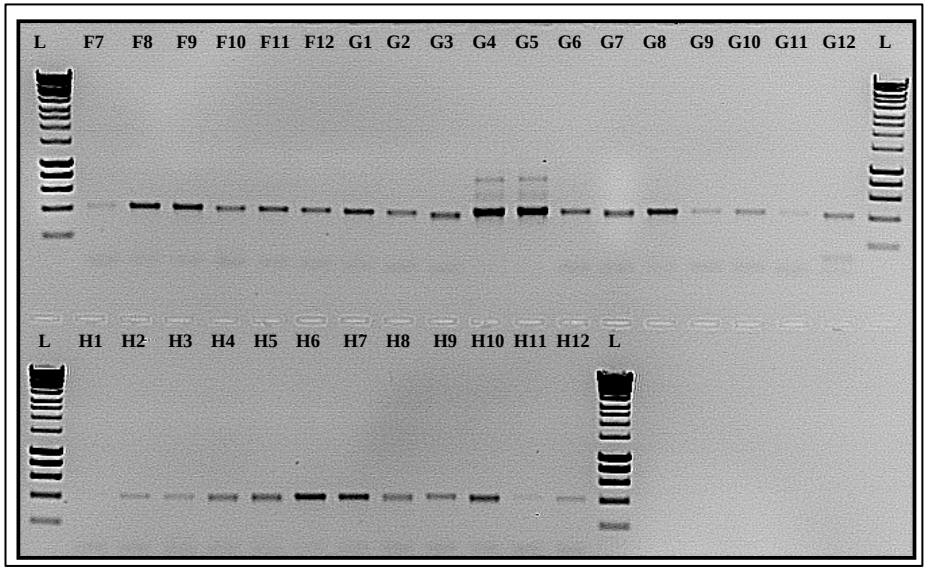


[44] The letter and number for each sample corresponds to the sample identification on the template for library 2 in Appendix 5. The '-' is the PCR negative control. Samples with weak amplification were repeated.

Library 3

27Frd1-338R-rd2 PCR results [45]

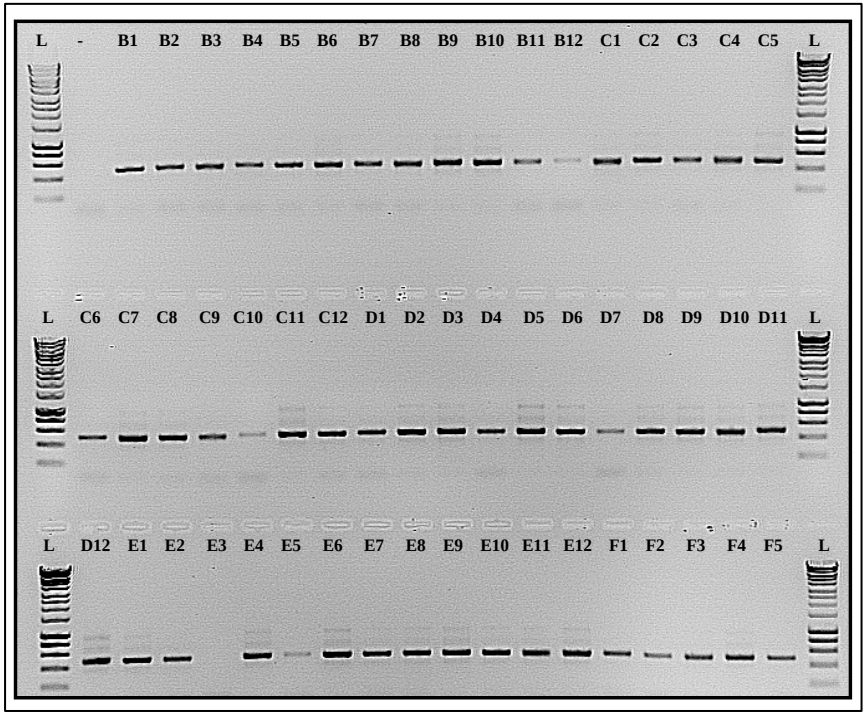


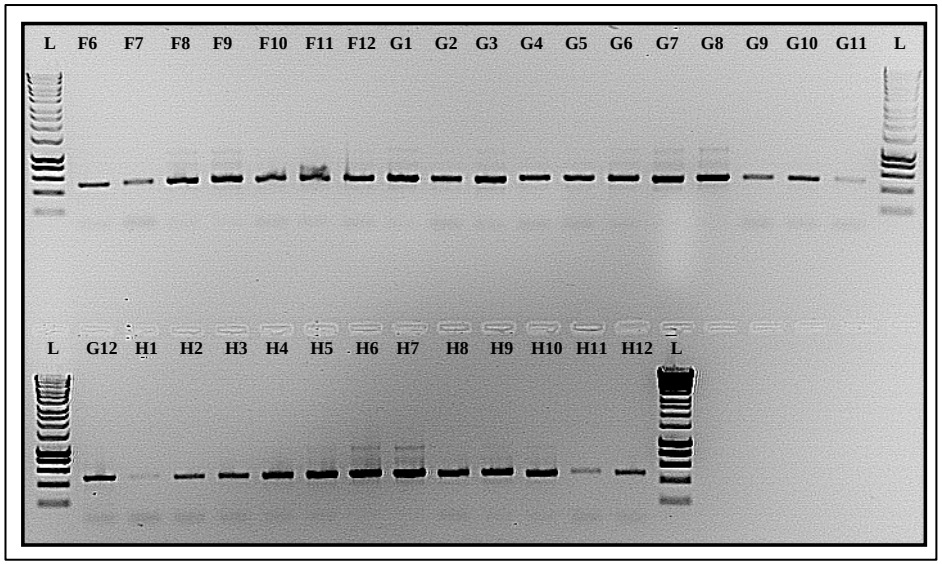


[45] The letter and number for each sample corresponds to the sample identification on the template for library 3 in Appendix 5. Samples with weak amplification were repeated.

Library 3

N5/N7 PCR results [46]

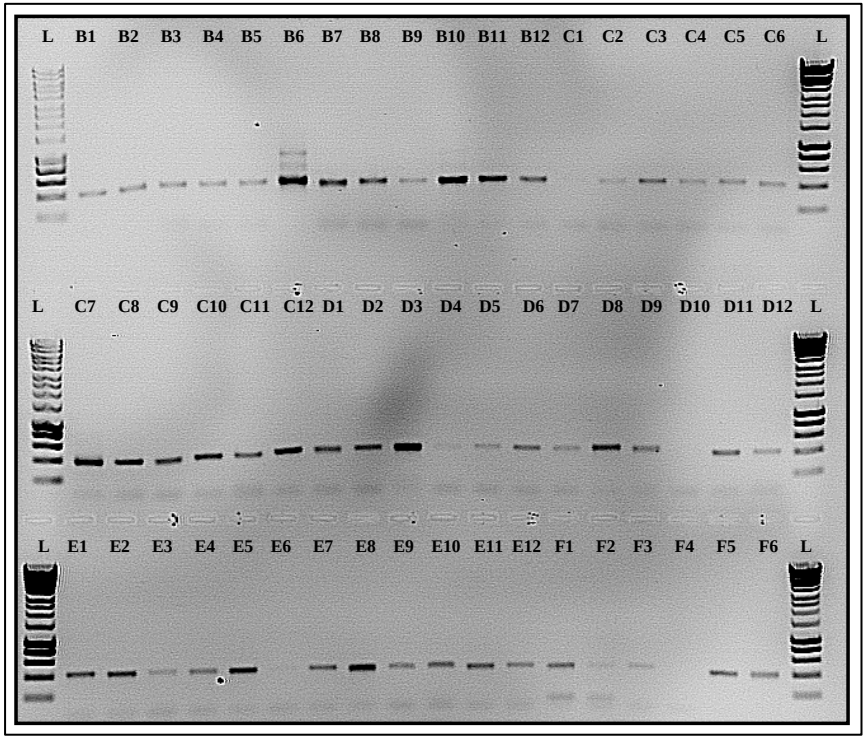


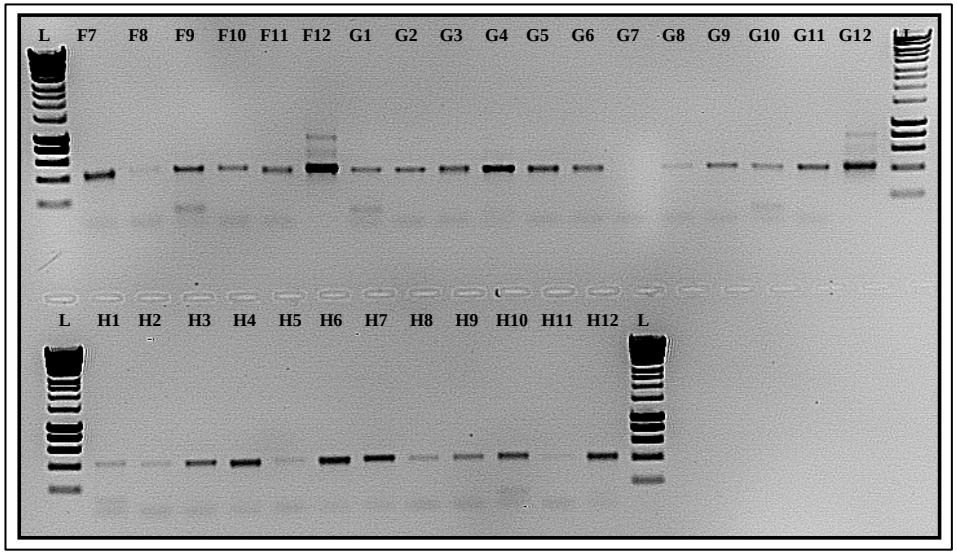


[46] The letter and number for each sample corresponds to the sample identification on the template for library 3 in Appendix 5. The '-' is the PCR negative control. Samples with weak amplification were repeated.

Library 4

27F-rd1/338R-rd2 PCR results [47]

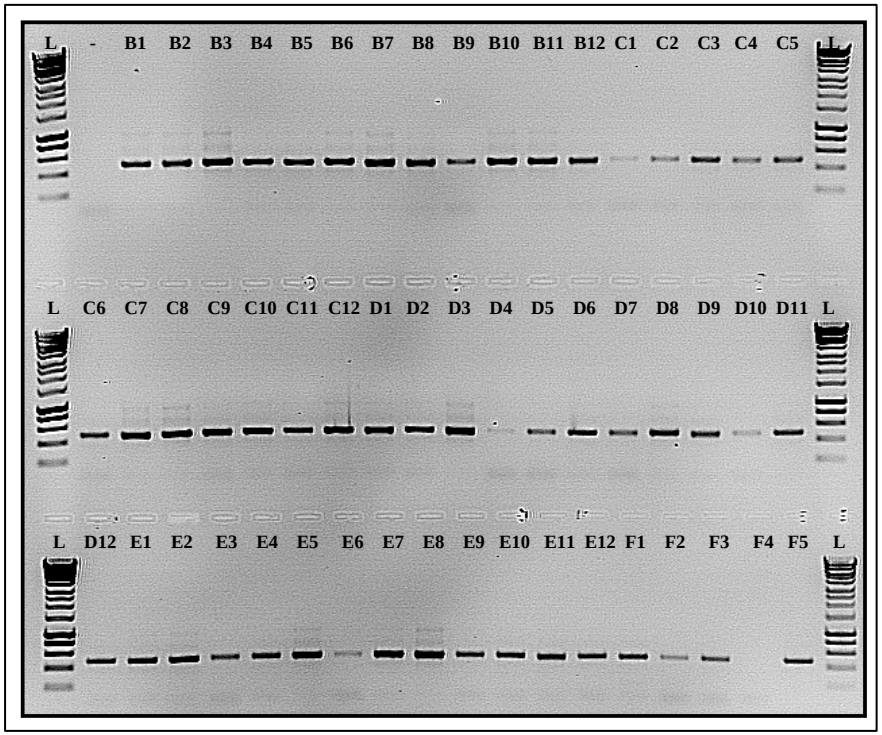


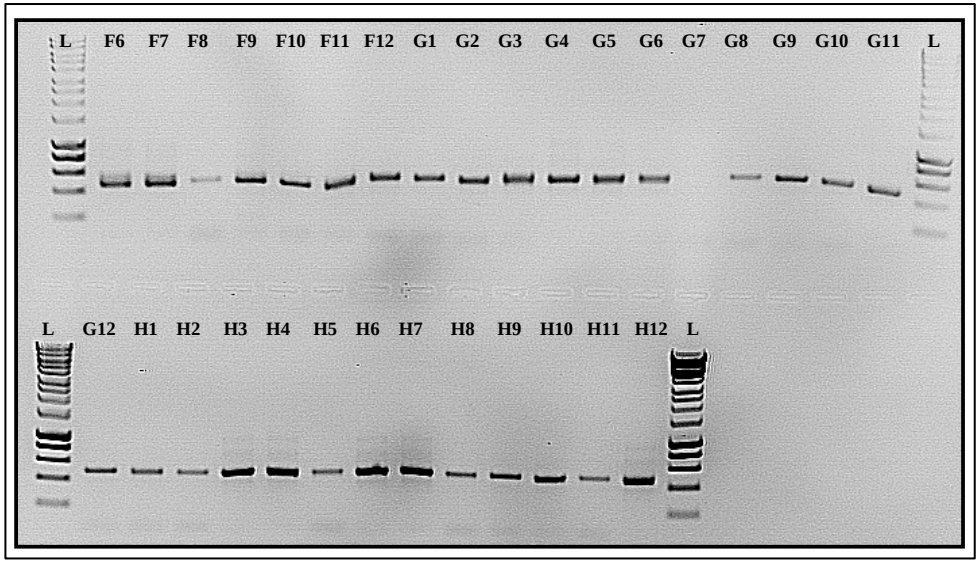


[47] The letter and number for each sample corresponds to the sample identification on the template for library 4 in Appendix 5. Samples with weak amplification were repeated.

Library 4

N5/N7 PCR results [48]

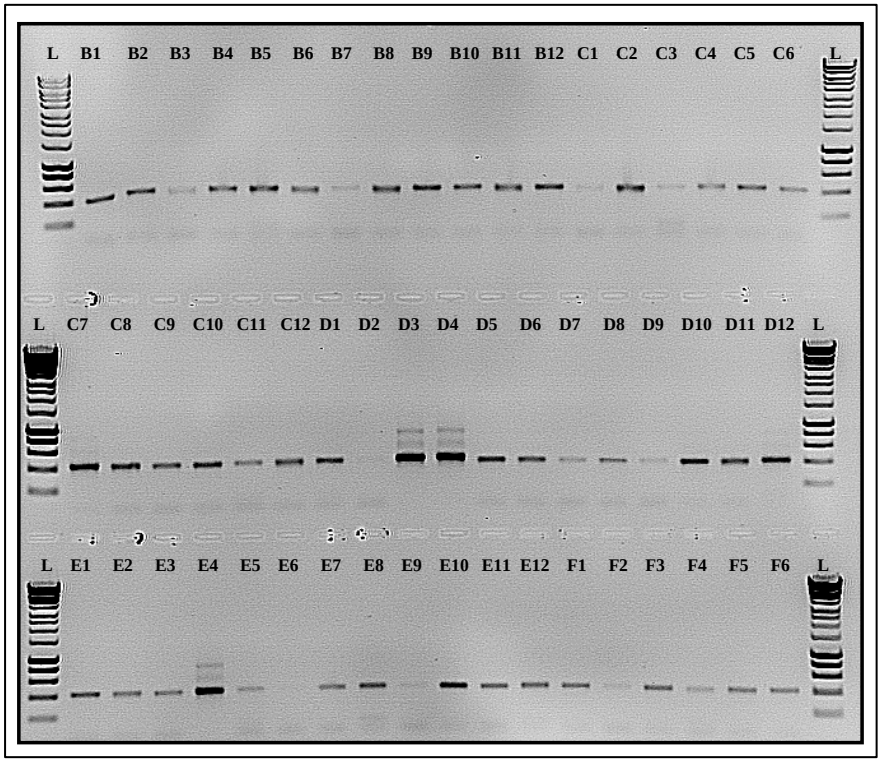


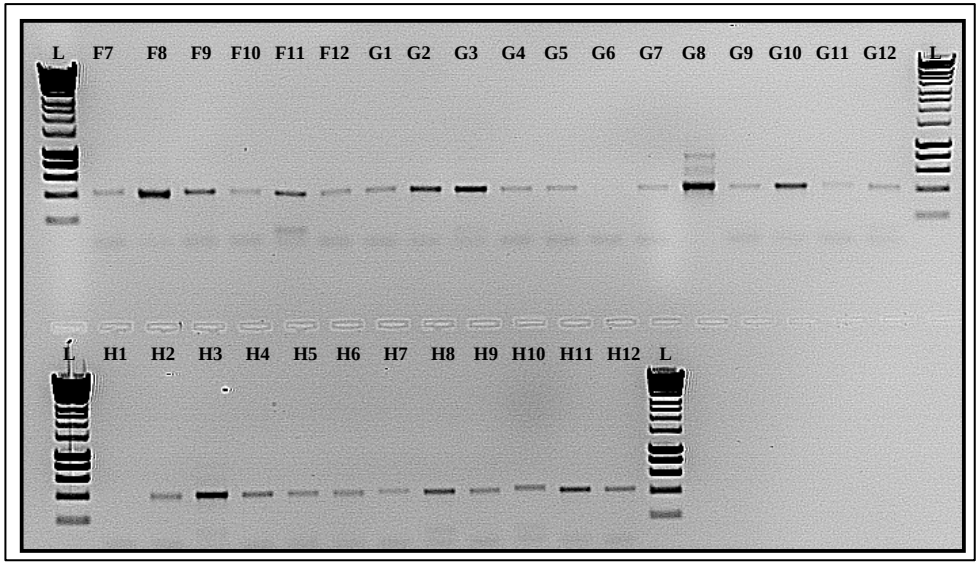


[48] The letter and number for each sample corresponds to the sample identification on the template for library 4 in Appendix 5. The '-' is the PCR negative control and G7 is a negative control carried through from the 27F-rd1/338R-rd2 PCR. Samples with weak amplification were repeated.

Library 5

27F-rd1/338R-rd2 PCR results [49]

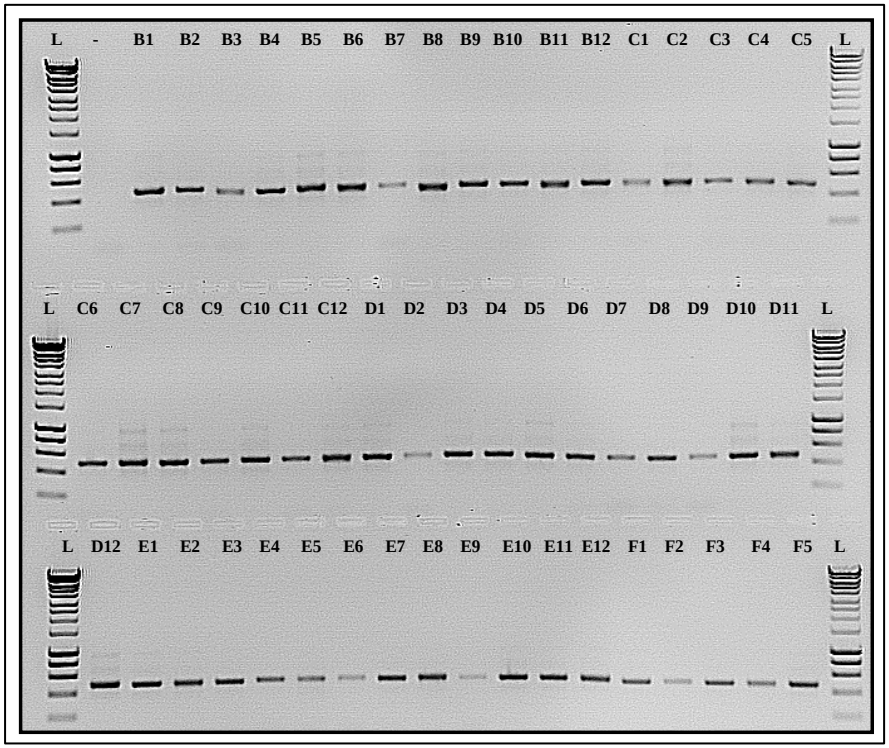


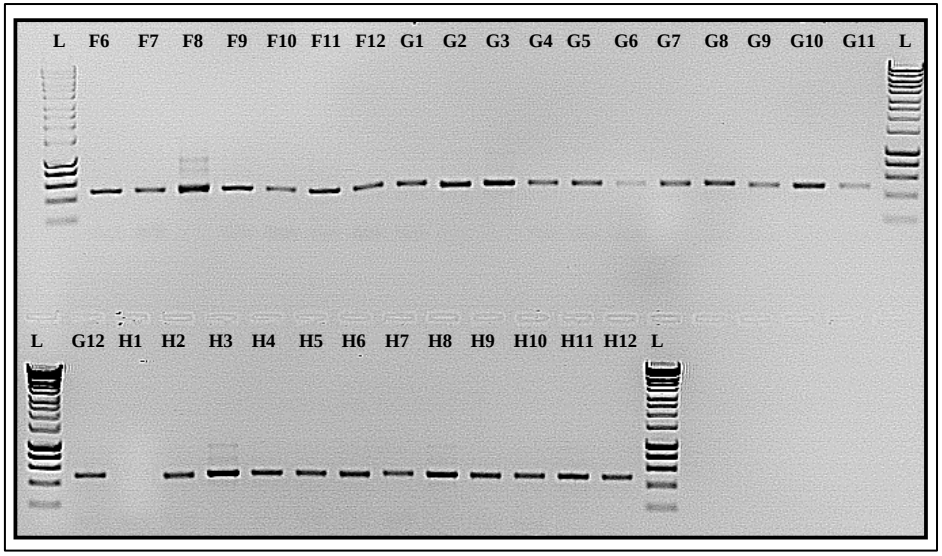


[49] The letter and number for each sample corresponds to the sample identification on the template for library 5 in Appendix 5. Samples with weak amplification were repeated.

Library 5

N5/N7 PCR results [50]





[50] The letter and number for each sample corresponds to the sample identification on the template for library 5 in Appendix 5. The '-' and H1 are PCR negative controls. Samples with weak amplification were repeated.